



TITLE SART-3 (Squamous cell carcinoma antigen recognized by T cells)  
JOURNAL Published Only in DataBase (1999)  
REFERENCE 2 (bases 1 to 3788)  
AUTHORS Itoh, K., Yang, D., Sasatomi, T., Nakao, M., Shichiho, S., Takasu, H.,  
Matsumoto, H., Mori, K. and Yamana, H.  
Direct Submission  
Submitted (07-DEC-1998) Kyogo Itoh, Kurume University, Department  
of Immunology; 67 Asahi-machi, Kurume, Fukuoka 830-0011, Japan  
(E-mail: kyogoi@med.kurume-u.ac.jp, Tel: 81-942-35-3311,  
Fax: 81-942-31-7699)  
FEATURES Location/Qualifiers  
source 1..3788  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/cell\_line="KE4"  
/cell\_type="squamous cell carcinoma"  
CDS 20..2911  
/note="squamous cell carcinoma antigen recognized by T  
cells"  
/codon\_start=1  
/product="squamous cell carcinoma antigen SART-3"  
/protein\_id="BAA78384.1"  
/db\_xref="GI:4996282"  
/translation="MATAEETASPEAESKAPKADGEEDEKVAARTRRKVLRSRAV  
AATYKMGPAWQOEGVSESDGEYAMASSPGEYENYDEEERKNOLEIERLE  
EQSLNHYNDYHVDLRLRLLEGLTKYRMARQKMSFIPTEELWLEWHLDEISMA  
QGLDREHYVDLFEKAVKDYICPNWLEKYQSVGKGLKRVSRFERALSSVG  
LHMTGLALWEAFREFSAIYEAARLEKVLHFRROLATPLDYDMEATFAEYEWSEDP  
IPESVIONYNKALQLEKVPYEEALLOAEPLAEOAYIDFENKIGDPARLOLIFE  
RALVENCPLDNLIRYSQYLDRLQKVDLVLVSHNRARINCPWTVALWRSYLLAMERH  
GVHQVISTFEKALNAGFIQATVVEIWOAYLDYLRVRVDFKQSSKELEELRAFT  
RALEYLUKVEERFENSGPSCVIMONWARIELRNKMKARELWDSIMTRGNAKYA  
NMWLEYNLERAHGDTQHCRLHRAVQTSYDPEHVCVELLTMERTESLESDWDIAY  
QKTEFLARNDQKRAEKAALVOOEEKAEORRAEKALKKKKIRGPEKRG  
ADDDKESGDEDEEOPSKRRRVENSIPAAETQNVVEAGPAGCAADVPEPSKQK  
EKAASLRDMPKVLHDSKDSITVFNLSNIPYSMQBPDTKRLPLFACRGVQIRPIFS  
NKGDFRGYCYFEKESKALQALEMDRKSVEGRPMFVSPVDKSNKDPKFRYRSTSL  
EKHLFISGLPFSCKEELEECAGHTVKGLDRLVNRAGKPKGLAYVYENESQSQ  
AVMKMDMTIKENIKVAISNPPQKVPKPKTRKAPGGPMLLPQTYGARGKRTQLS  
LLPLRALRPSAAPQAENGPAAPAAVAAPAAEAPAKMSNADEAKFLRK"  
polyA\_site 3788  
/note="18 A nucleotides"  
BASE COUNT 1020 a 846 c 1081 g 841 t  
ORIGIN  
Query Match 99.3%; Score 3771.2; DB 9; Length 3788;  
Best Local Similarity 99.9%; Pred. No. 0;  
Matches 3773; Conservative 0; Mismatches 3; Indels 0; Gaps 0;  
QY 5 GGTCCGATGGCGACTGCGGCGGAACCTCGGCTTCAGAACCGGAGGCTGAGTCCAAAGC 64  
DB 13 GCGCAAGATGGCGACTGCGGCGGAACCTCGGCTTCAGAACCGGAGGCTGAGTCCAAAGC 72  
QY 65 TGGGCGCAAGGCTGACGAGAGAGGATGAGGTAAAGCGGCTAGCACAAAGGAAAGCT 124  
DB 73 TGGGCGCAAGGCTGACGAGAGAGGATGAGGTAAAGCGGCTAGCACAAAGGAAAGCT 132  
QY 125 GTTATCGCGGCTGTGCGCGCTCGCACATACAGACCATGGGCGCAGCGCTGGGATCAGCA 184  
DB 133 GTTATCGCGGCTGTGCGCGCTCGCACATACAGACCATGGGCGCAGCGCTGGGATCAGCA 192  
QY 185 GGAGGAAGCGGTGACGAGAGGATGGGATGAGTACGCCATCGGCTTCCTCCGCGGAGAG 244  
DB 193 GGAGGAAGCGGTGACGAGAGGATGGGATGAGTACGCCATCGGCTTCCTCCGCGGAGAG 252  
QY 245 CTCCTCCCGGGAGTACGAGTGGGAATATACCAAGAGGAGGAGAAACACGCTGGAGAT 304  
DB 253 CTCCTCCCGGGAGTACGAGTGGGAATATACCAAGAGGAGGAGAAACACGCTGGAGAT 312  
QY 305 TGAGAGACTGGAGGAGCACTTGTCTATCAACGCTATGACTACCAACTGCCATGTGGACTT 364  
DB 313 TGAGAGACTGGAGGAGCACTTGTCTATCAACGCTATGACTACCAACTGCCATGTGGACTT 372

Db 1453 GCTAGGATGAGGCTGACTGTCATTAACATGAGAAAGCTCGGAACCTCTGGGATAG 1512  
QY 1505 CATCATGACGAGAGAAATGCCAGTACGCCCAACATGTGGCTAGAGTATTACAACCTGGA 1564  
Db 1513 CATCATGACGAGAGAAATGCCAGTACGCCCAACATGTGGCTAGAGTATTACAACCTGGA 1572  
QY 1565 AAGAGCTCATGGTGACACCCAGCACTGCGGGAAGGCTCTGCACCGGGCCGTCAGATGCAC 1624  
Db 1573 AAGAGCTCATGGTGACACCCAGCACTGCGGGAAGGCTCTGCACCGGGCCGTCAGATGCAC 1632  
QY 1625 CAGTACTACCCAGAGACAGTCTGCGAAGTGTACTACCATGAGAGAGAGAGAGGTTTC 1684  
Db 1633 CAGTACTACCCAGAGACAGTCTGCGAAGTGTACTACCATGAGAGAGAGAGGTTTC 1692  
QY 1685 TTTAGAAGATTGGGATATAGCTGTTTCAGAAATCGAAACCCGATTAGCTCGTCAATGA 1744  
Db 1693 TTTAGAAGATTGGGATATAGCTGTTTCAGAAATCGAAACCCGATTAGCTCGTCAATGA 1752  
QY 1745 GCAGAGATGAAGGCTGCAG 1804  
Db 1753 GCAGAGATGAAGGCTGCAG 1812  
QY 1805 TGAACAACGAG 1864  
Db 1813 TGAACAACGAG 1872  
QY 1865 AGGCCACAG 1924  
Db 1873 AGGCCACAG 1932  
QY 1925 AGAGCAGGCTTCCAAACGAG 1984  
Db 1933 AGAGCAGGCTTCCAAACGAG 1992  
QY 1985 AAATGTAGAGTAGCAG 2044  
Db 1993 AAATGTAGAGTAGCAG 2052  
QY 2045 TTTGAG 2104  
Db 2053 TTTGAG 2112  
QY 2105 CAGCAG 2164  
Db 2113 CAGCAG 2172  
QY 2165 GGACAG 2224  
Db 2173 GGACAG 2232  
QY 2225 CTTGAGCAACCGTGGGATTTCCGAGGTTTACTGCTACGTGGAGTTTAAAGAGAGAAATC 2284  
Db 2233 CTTGAGCAACCGTGGGATTTCCGAGGTTTACTGCTACGTGGAGTTTAAAGAGAGAAATC 2292  
QY 2285 AGCCCTTCAGGCACTGGAG 2344  
Db 2293 AGCCCTTCAGGCACTGGAG 2352  
QY 2345 CCCCTGTGTGATAG 2404  
Db 2353 CCCCTGTGTGATAG 2412  
QY 2405 AGAGAAACAGAGCTGTTTCATCTCAGGCTGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 2464  
Db 2413 AGAGAAACAGAGCTGTTTCATCTCAGGCTGCTCTCTCTCTCTCTCTCTCTCTCTCTCT 2472  
QY 2465 AGAAATCTGAAGGCTCATGGCAGCCGTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2524  
Db 2473 AGAAATCTGAAGGCTCATGGCAGCCGTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2532  
QY 2525 CAAACCAAGGCGCTGAGTGTGAGAGATGAATGAATGAATGAATGAATGAATGAATGAATGAAT 2584  
|||||

Db 2533 CAAACCAAGGCGCTGAGTGTGAGAGATGAATGAATGAATGAATGAATGAATGAATGAATGAAT 2592  
QY 2585 GATGAAGATGGAGCGCATGACTATCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2644  
Db 2593 GATGAAGATGGAGCGCATGACTATCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2652  
QY 2645 TCCTCAGAGGAAAGTTCCAG 2704  
Db 2653 TCCTCAGAGGAAAGTTCCAG 2712  
QY 2705 TTTGCCGCGACATACGAG 2764  
Db 2713 TTTGCCGCGACATACGAG 2772  
QY 2765 TGCCCTCAGAGCGCCCAAGTGTGCTCAGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2824  
Db 2773 TGCCCTCAGAGCGCCCAAGTGTGCTCAGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2832  
QY 2825 TGCAGTTGCCGCGCCCAAGTGTGCTCAGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2884  
Db 2833 TGCAGTTGCCGCGCCCAAGTGTGCTCAGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2892  
QY 2885 GCTGTTCTGAG 2944  
Db 2893 GCTGTTCTGAG 2952  
QY 2945 CCGCGGAGAGCTCCCAAGTGTGCTCAGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3004  
Db 2953 CCGCGGAGAGCTCCCAAGTGTGCTCAGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3012  
QY 3005 TGCTCGCAACCAAGTGTGCTCAGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3064  
Db 3013 TGCTCGCAACCAAGTGTGCTCAGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3072  
QY 3065 AGAGCTTTTCAGTGTCTCTATATGAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3124  
Db 3073 AGAGCTTTTCAGTGTCTCTATATGAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3132  
QY 3125 CTGGGCGCAGAGTGTGCTCTGAGATAGTTTGTATCTTAAAGAGAGAGAGAGAGAGAGAGAG 3184  
Db 3133 CTGGGCGCAGAGTGTGCTCTGAGATAGTTTGTATCTTAAAGAGAGAGAGAGAGAGAGAGAG 3192  
QY 3185 CGAAACGAGAGACACACTGTTTGTGAGACAGTGTGCTCAATGTTTCTGGCAGCTCTCCG 3244  
Db 3193 CGAAACGAGAGACACACTGTTTGTGAGACAGTGTGCTCAATGTTTCTGGCAGCTCTCCG 3252  
QY 3245 CCCCTTTTGTATGACACTTCTTCCACCTGACAGACAGTGTGCTCAATGTTTCTGGCAGCTCT 3304  
Db 3253 CCCCTTTTGTATGACACTTCTTCCACCTGACAGACAGTGTGCTCAATGTTTCTGGCAGCTCT 3312  
QY 3305 AATTTTAAAGATGAAATGGCAGAGTGTAGTAAATTCACAGAGTGTGCTCTGTTGGGGTG 3364  
Db 3313 AATTTTAAAGATGAAATGGCAGAGTGTAGTAAATTCACAGAGTGTGCTCTGTTGGGGTG 3372  
QY 3365 GGTCTGAGGAGAGTCAAGCTATAAACAATTTGCTGGAGTTTGTTCATGAGGCTGTGTCAT 3424  
Db 3373 GGTCTGAGGAGAGTCAAGCTATAAACAATTTGCTGGAGTTTGTTCATGAGGCTGTGTCAT 3432  
QY 3425 TTTTATATATGTTGTTGTAATGACATGACAGCTTCTTTCATGTTTCTTAAAGAGAGAG 3484  
Db 3433 TTTTATATATGTTGTTGTAATGACATGACAGCTTCTTTCATGTTTCTTAAAGAGAGAG 3492  
QY 3485 AATATTGCAACATTTGTTTGTATAGGAATTTATTTGTGCCACCTCTGCTGGAGCTGTTT 3544  
Db 3493 AATATTGCAACATTTGTTTGTATAGGAATTTATTTGTGCCACCTCTGCTGGAGCTGTTT 3552  
QY 3545 CTTTGCCTAGTACAGTACCTGTGTTCTTAAACATGAGTTTCAAGCTTTTGGGTTTGG 3604  
Db 3553 CTTTGCCTAGTACAGTACCTGTGTTCTTAAACATGAGTTTCAAGCTTTTGGGTTTGG 3612  
QY 3605 TTTAATACCATGTCAAAATGCAAACTCAATTTCCCCCATTTAGCTTTTATTAACACTGACGT 3664  
Db 3613 TTTAATACCATGTCAAAATGCAAACTCAATTTCCCCCATTTAGCTTTTATTAACACTGACGT 3672  
|||||

QY 3665 TCTCTTCAAACTTCTTCTGTAATGCTACTAGATGTGCTATTCACATACAGATGTGTTT 3724  
|||||  
Db 3673 TCTCTTCAAACTTCTTCTGTAATGCTACTAGATGTGCTATTCACATACAGATGTGTTT 3732  
|||||  
QY 3725 GAAGTGGGTGACCTTCTTACCTAATAGATGTGTAATAGAACTTTTGTAGTC 3780  
|||||  
Db 3733 GAAGTGGGTGACCTTCTTACCTAATAGATGTGTAATAGAACTTTTGTAGTC 3788  
|||||

## RESULT 2

LOCUS D63879 3660 bp mRNA linear PRI 06-OCT-2001  
DEFINITION Human mRNA for KIAA0156 gene, complete cds.  
ACCESSION D63879  
VERSION D63879.1 GI:961449  
KEYWORDS KIAA0156.  
SOURCE Homo sapiens male myeloblast cell\_line:KG-1 CDNA to mRNA.  
ORGANISM Homo sapiens

REFERENCE 1 Nagase, T., Seki, N., Tanaka, A., Ishikawa, K. and Nomura, N.  
AUTHORS Prediction of the coding sequences of unidentified human genes. IV.  
TITLE The coding sequences of 40 new genes (KIAA0121-KIAA0160) deduced by  
analysis of cDNA clones from human cell line KG-1  
JOURNAL DNA Res. 2 (4), 167-174 (1995)  
MEDLINE 96127530  
AUTHORS 2 (bases 1 to 3660)  
TITLE Ohara, O., Nagase, T., Kikuno, R. and Nomura, N.  
JOURNAL Direct Submission

REFERENCE 1 Nagase, T., Seki, N., Tanaka, A., Ishikawa, K. and Nomura, N.  
AUTHORS Prediction of the coding sequences of unidentified human genes. IV.  
TITLE The coding sequences of 40 new genes (KIAA0121-KIAA0160) deduced by  
analysis of cDNA clones from human cell line KG-1  
JOURNAL DNA Res. 2 (4), 167-174 (1995)  
MEDLINE 96127530  
AUTHORS 2 (bases 1 to 3660)  
TITLE Ohara, O., Nagase, T., Kikuno, R. and Nomura, N.  
JOURNAL Direct Submission

## FEATURES

Source  
1..3660  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/sex="male"  
/cell\_line="KG-1"  
/cell\_type="myeloblast"  
1..3660  
/gene="KIAA0156"  
1..2892  
/gene="KIAA0156"  
/note="KIAA0156 gene product is related to xenopus  
nucleolin."  
/codon\_start=1  
/protein\_id="BAA09929.1"  
/db\_xref="GI:961450"

/translation="MATAETTSASEPEAESKAGRADGEDEEVKAARTRRKVLRAVA  
AATYTMGPAMDQEGYSESDGEYAMASSAESSPGYEWYDEEEKNOLEIERLE  
EQLSNVDYNDCHVDLIRLLLEGELTKVRMAROKMSEIFPLETELEWLHDEISMA  
ODGLREHYDLFEKAVDYICPNWLEYGYQSVGIGQKGLKRVSVFERALSSVG  
LHMTKGLALWEYREFSAIPEARLEKVSFLRQLAIPLYDMEATAEYEWESEDP  
IPESVINYNAKLOQERKYPERALLOAEAPRAEYQAYIDFENKIDGPARIQIIFE  
RALVENCPLDLMIRYSOYLDRQLKDLVLSVHNRAIRNCPWTVALWSRYLLAMERH  
GYDHOVISVTFPEKALNAGFIQATDYVETWQYLDYLRVRDFKQDSKELELRAAF  
RALEYLNLERHREFNESGDPSCVINONNARIEARLCNNOKARELWDSITMTRGNKAYA  
QKTEFLRANRFORMAAEKALVQEEERAEQRRKAREKALKKKKIKGPEKRG  
ADEDEKEMGDDEEQPKRRRVENSIPAAETONVEAAGPAGCAADVPPSPKQK  
EKAASLKRDMPKLVHDSKDSITVFNPLPYSMQPDTKLRPLFEACQVQIRFIS  
NRGDFRGYCYVEFEKESALQALEMDKSVGRPMFVSPVDKSNPDPKVFYRSTSL  
EKHKLFISGLPEFCEKLEETICKAHGTVDLRLVTRNAGPKGLAYVEYENESQAO  
AVMKDGTIKENIKVAISNPPOKRVPEKPEKTRKAPGPMPLLPOTYGARGKRTQLS  
LPLRALQPSAAAPQENGPAAPAAVAAPAAPEAPKMSNADEFAKFLRK"

BASE COUNT 986 a 824 c 1049 g 801 t  
ORIGIN  
Query Match 96.3%; Score 3656.8; DB 9; Length 3660;  
Best Local Similarity 99.9%; Pred. No. 0;  
Matches 3658; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 12 ATGGGACTGCGCGCGAAACCTCGGCTTCAGAACCCGAGGCTAGTCCAAAGGCTGGGCC 71  
|||||  
Db 1 ATGGGACTGCGCGCGAAACCTCGGCTTCAGAACCCGAGGCTAGTCCAAAGGCTGGGCC 60  
|||||  
QY 72 AAGGCTGACGGAGAGAGGATGAGTTAAGCGGCTAGGACAGGAGAGAGGTTATCG 131  
|||||  
Db 61 AAGGCTGACGGAGAGAGGATGAGTTAAGCGGCTAGGACAGGAGAGGTTATCG 120  
|||||  
QY 132 CGGGCTGTGGCGCTGCGACATACAGACCATAGGGCCAGCGTGGGATCAGCAGAGGAA 191  
|||||  
Db 121 CGGGCTGTGGCGCTGCGACATACAGACCATAGGGCCAGCGTGGGATCAGCAGAGGAA 180  
|||||  
QY 192 GCGCTGACGGAGAGAGGATGGGATGAGTACGCCATGGCTTCCTCCGCGAGAGCTCC 251  
|||||  
Db 181 GCGCTGACGGAGAGAGGATGGGATGAGTACGCCATGGCTTCCTCCGCGAGAGCTCC 240  
|||||  
QY 252 GGGGAGTACGAGTGGGATATACGAGAGGAGGAGAGGAGGAGGAGGAGGAGGAGGAG 311  
|||||  
Db 241 GGGGAGTACGAGTGGGATATACGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 300  
|||||  
QY 312 CTGGAGGAGCAGTGTCTATCAACGCTATGACTACAACTGCGCATGTGACTTGATCAGA 371  
|||||  
Db 301 CTGGAGGAGCAGTGTCTATCAACGCTATGACTACAACTGCGCATGTGACTTGATCAGA 360  
|||||  
QY 372 CTGCTCAGGCTGGAAGGGAGGCTTACCAAGGTGAGGATGCGCGCGCAGAGATGACTGAA 431  
|||||  
Db 361 CTGCTCAGGCTGGAAGGGAGGCTTACCAAGGTGAGGATGCGCGCGCAGAGATGACTGAA 420  
|||||  
QY 432 ATCTTTCCCTTACCTGAAGAGCTCTGGCTGGAGTGGCTGCATGACAGATCAGCATGGCC 491  
|||||  
Db 421 ATCTTTCCCTTACCTGAAGAGCTCTGGCTGGAGTGGCTGCATGACAGATCAGCATGGCC 480  
|||||  
QY 492 CAGGATGGCTGGACAGAGAGCAGCTGTATGACCTCTTTGAGAAAGCCCTGAGGATTA 551  
|||||  
Db 481 CAGGATGGCTGGACAGAGAGCAGCTGTATGACCTCTTTGAGAAAGCCCTGAGGATTA 540  
|||||  
QY 552 ATTTGCTTACATTTGGCTAGAGTATGCCAGTACTCATGTTGGTGGGATTTGGTCAGAA 611  
|||||  
Db 541 ATTTGCTTACATTTGGCTAGAGTATGCCAGTACTCATGTTGGTGGGATTTGGTCAGAA 600  
|||||  
QY 612 GGTGGCTTGAAAGTTGCTCGCTGGTTGAAAGGCTCTCTGCTCTGTTGGTTTACAT 671  
|||||  
Db 601 GGTGGCTTGAAAGTTGCTCGCTGGTTGAAAGGCTCTCTGCTCTGTTGGTTTACAT 660  
|||||  
QY 672 ATGACCAAGAGCTCGCCCTCTGGGAGGCTTACCGAGAGTTCGAAAGTGCAGTTGTGAA 731  
|||||  
Db 661 ATGACCAAGAGCTCGCCCTCTGGGAGGCTTACCGAGAGTTCGAAAGTGCAGTTGTGAA 720  
|||||  
QY 732 GCTGCTCGGCTTGAAAGTTCACAGTCTTTTCCGGCGAGTGGCGATCCCACTCTAT 791  
|||||  
Db 721 GCTGCTCGGCTTGAAAGTTCACAGTCTTTTCCGGCGAGTGGCGATCCCACTCTAT 780  
|||||  
QY 792 GATATGGAGCCACATTTGCAGAGTATGAAGATGGTTCAGAGACCCATACAGAGTCA 851  
|||||  
Db 781 GATATGGAGCCACATTTGCAGAGTATGAAGATGGTTCAGAGACCCATACAGAGTCA 840  
|||||  
QY 852 GTAATTCAAGAACTATAAACAAGCACTACAGAGCTGGAGAAATATAAACCCTATGAAGAA 911  
|||||  
Db 841 GTAATTCAAGAACTATAAACAAGCACTACAGAGCTGGAGAAATATAAACCCTATGAAGAA 900  
|||||  
QY 912 GCACCTGTTGACGAGAGAGCACCAGGCTGGCAGAAATATCAAGCATATATCGATTTTGGAG 971  
|||||  
Db 901 GCACCTGTTGACGAGAGAGCACCAGGCTGGCAGAAATATCAAGCATATATCGATTTTGGAG 960  
|||||  
QY 972 ATGAAATTTGGGATCTGCTCGCATTCAGTTGATCTTTGAGCGGCCCTGTCGAGAAC 1031  
|||||  
Db 961 ATGAAATTTGGGATCTGCTCGCATTCAGTTGATCTTTGAGCGGCCCTGTCGAGAAC 1020  
|||||  
QY 1032 TGCCTTTGCCAGACTTATGATTCGTTACAGTACAGTACAGTACAGTACAGTACAGTACAGTAC 1091  
|||||  
Db 1021 TGCCTTTGCCAGACTTATGATTCGTTACAGTACAGTACAGTACAGTACAGTACAGTACAGTAC 1080  
|||||



QY	1092	AAGGATTGGTTTATCTGTACATAACCGCGCTATTAGAACTGCCCTGGACAGTTGCC	1151		2161	AACTCAGGGCACTCTCTCGAGGCTGTGGGAGGTGGTCAGATCCGACCATCTTCAGC	2220	
Db	1081	AAGGATTGGTTTATCTGTACATAACCGCGCTATTAGAACTGCCCTGGACAGTTGCC	1140		2232	AACCGTGGGATTTCCGAGTTTACTGCTAGCTGGAGTTTAAAGAGAAATCAGCCCTT	2291	
QY	1152	TTATGAGTGGTACCTCTTGGCCATGGAGACATGGAGTTGATCATCAAGTAATTTCT	1211		2221	AACCGTGGGATTTCCGAGTTTACTGCTAGCTGGAGTTTAAAGAGAAATCAGCCCTT	2280	
Db	1141	TTATGAGTGGTACCTCTTGGCCATGGAGACATGGAGTTGATCATCAAGTAATTTCT	1200		2292	CAGGCACTGGAGATGGACCGAAAAGTGTAGAGGGAGGCAATGTTTGTTCCTCCCTGT	2351	
QY	1212	GTAACTTTCAGAAAGCTTTGAATGCCGGCTTCCATCCAGGCCACTGATTATGTGGAGATT	1271		2281	CAGGCACTGGAGATGGACCGAAAAGTGTAGAGGGAGGCAATGTTTGTTCCTCCCTGT	2340	
Db	1201	GTAACTTTCAGAAAGCTTTGAATGCCGGCTTCCATCCAGGCCACTGATTATGTGGAGATT	1260		2352	GTGGATAAGAGCAAAAAACCCCGATTTTAAGGTGTTTACAGTACAGCACTTCCCTAGAGAAA	2411	
QY	1272	TGGAGGCAATACCTGATTACCTTGAGGAGAAAGGTTGATTCAAAACAAGACTCCAGTAAA	1331		2341	GTGGATAAGAGCAAAAAACCCCGATTTTAAGGTGTTTACAGTACAGCACTTCCCTAGAGAAA	2400	
Db	1261	TGGAGGCAATACCTGATTACCTTGAGGAGAAAGGTTGATTCAAAACAAGACTCCAGTAAA	1320		2412	CACAAGCTGTTCATCTCAGGCCCTGCCCTTCTCTCTGTACTTAAAGAGAACTAGAGAAATC	2471	
QY	1332	GAGCTGGAGAGTTGAGGGCCGCTTTACTCGTGGCTTGAGTATCTGAAGCAGGAGGTG	1391		2401	CACAAGCTGTTCATCTCAGGCCCTGCCCTTCTCTCTGTACTTAAAGAGAACTAGAGAAATC	2460	
Db	1321	GAGCTGGAGAGTTGAGGGCCGCTTTACTCGTGGCTTGAGTATCTGAAGCAGGAGGTG	1380		2472	TGTAAGGCTCATGCGACCGGTGAAGGACCTCAGGCTGTGTACCAACCCGGCTGGCAACCA	2531	
QY	1392	GAAGAGCTTTCATGAGAGTGTGTATCCAAAGCTGGGTGATTATGAGAACTGGGTAGG	1451		2461	TGTAAGGCTCATGCGACCGGTGAAGGACCTCAGGCTGTGTACCAACCCGGCTGGCAACCA	2520	
Db	1381	GAAGAGCTTTCATGAGAGTGTGTATCCAAAGCTGGGTGATTATGAGAACTGGGTAGG	1440		2532	AAGGGCTGGCTTACGTGGAGTATGAAAATGAATCCAGGGCTGCGAGGCTGTGATGAAG	2591	
QY	1452	ATTGAGGCTCGATGTGCAATAACATGCAGAAAGCTCGGAACTCTGGGATAGCATATG	1511		2521	AAGGGCTGGCTTACGTGGAGTATGAAAATGAATCCAGGGCTGCGAGGCTGTGATGAAG	2580	
Db	1441	ATTGAGGCTCGATGTGCAATAACATGCAGAAAGCTCGGAACTCTGGGATAGCATATG	1500		2592	ATGAGCGGCATGACTATCAAGAGAACATCATCAAGTGGCAATCAGCAACCCCTCCCTCAG	2651	
QY	1512	ACCAGAGGAATGCCAAGTACGCCAATGTGCTAGAGTATTACAACTGGAAAGAGCT	1571		2581	ATGAGCGGCATGACTATCAAGAGAACATCATCAAGTGGCAATCAGCAACCCCTCCCTCAG	2640	
Db	1501	ACCAGAGGAATGCCAAGTACGCCAATGTGCTAGAGTATTACAACTGGAAAGAGCT	1560		2652	AGAAAGTTCCAGAGAAAGCAGAGACCCAGGAAGCACCAGTGGCCCTATGCTTTTGCCG	2711	
QY	1572	CATGCTGACACCCAGCACTGCCGGAAGGCTCTGCACCGGCCCTCCAGTGCACCACTGAC	1631		2641	AGAAAGTTCCAGAGAAAGCAGAGACCCAGGAAGCACCAGTGGCCCTATGCTTTTGCCG	2700	
Db	1561	CATGCTGACACCCAGCACTGCCGGAAGGCTCTGCACCGGCCCTCCAGTGCACCACTGAC	1620		2712	CAGACATACGAGGAGGGGAAAGGAGACGACGCTGTCTTACTGCTCTGTCGCTG	2771	
QY	1632	TACCCAGAGCACTCTGCGAAGTGTACTCACCATGGAGAGGACAGAAAGTCTTTAGAA	1691		2701	CAGACATACGAGGAGGGGAAAGGAGACGACGCTGTCTTACTGCTCTGTCGCTG	2760	
Db	1621	TACCCAGAGCACTCTGCGAAGTGTACTCACCATGGAGAGGACAGAAAGTCTTTAGAA	1680		2772	CAGGCCCCAAGTCTGCAGCTCCTCAGGCTGAGAAACCGCCCTGCCGCGCTCCCTGACGTT	2831	
QY	1692	GATTGGATATAGCTGTTCAGAAACTGAAACCCGATTAGCTGTCTCAATGACGAGAGA	1751		2761	CAGGCCCCAAGTCTGCAGCTCCTCAGGCTGAGAAACCGCCCTGCCGCGCTCCCTGACGTT	2820	
Db	1681	GATTGGATATAGCTGTTCAGAAACTGAAACCCGATTAGCTGTCTCAATGACGAGAGA	1740		2832	GCGCCCCAGCAGCCAGGACCCCAAGATGTCCCAATGCCGATTTTGGCAAGCTGTTT	2891	
QY	1752	ATGAAGGCTCAGAGAGGAAGCAGCCCTTGTGCAGCAAGAAAGAAAGGCTGACAA	1811		2821	GCGCCCCAGCAGCCAGGACCCCAAGATGTCCCAATGCCGATTTTGGCAAGCTGTTT	2880	
Db	1741	ATGAAGGCTCAGAGAGGAAGCAGCCCTTGTGCAGCAAGAAAGAAAGGCTGACAA	1800		2892	CTGAGAAAGTGAACGGGACGCTGGGAGACAGGAAATGCCCTTACTTCTGCTGCGCCGCG	2951	
QY	1812	CGAAAGAGCTCGGGCTGAGAGAAAGCGTTTAAAAAGAGAAAGATCAGAGGCCCA	1871		2881	CTGAGAAAGTGAACGGGACGCTGGGAGACAGGAAATGCCCTTACTTCTGCTGCGCCGCG	2940	
Db	1801	CGAAAGAGCTCGGGCTGAGAGAAAGCGTTTAAAAAGAGAAAGATCAGAGGCCCA	1860		2952	GACCTCCACACCCAGCAGTGCATGGGATGGAGAGGCTGGTGTCTGCTGCTGCTGCTG	3011	
QY	1872	GAGAGCGCGGACAGATGAGGACGATGAGAAAGATGGGGCGATGATGAAGAGAGCAG	1931		2941	GACCTCCACACCCAGCAGTGCATGGGATGGAGAGGCTGGTGTCTGCTGCTGCTGCTG	3000	
Db	1861	GAGAGCGCGGACAGATGAGGACGATGAGAAAGATGGGGCGATGATGAAGAGAGCAG	1920		3012	AACACAGATGGCTCTCGGCTTTAGACAGAAAGGGGAAAGGGTTCTTAAGTCAAGAGCCT	3071	
QY	1932	CCTTCCAAAGCAGAGGGTGCAGAACAGCATCCCTGCAGCTGGAGAAACACAAAATGTA	1991		3001	AACACAGATGGCTCTCGGCTTTAGACAGAAAGGGGAAAGGGTTCTTAAGTCAAGAGCCT	3060	
Db	1921	CCTTCCAAAGCAGAGGGTGCAGAACAGCATCCCTGCAGCTGGAGAAACACAAAATGTA	1980		3072	TTGAGTGTCTCTCATATGAGGCGAGTGGCAGAAAAGTGAACCACTCTGAGGCTGGGCC	3131	
QY	1992	GAAGTACGACAGGGCCGCTGGGAAATGTGCTGCCGTAGATGTGGAGCCCTTCGAAG	2051		3061	TTGAGTGTCTCTCATATGAGGCGAGTGGCAGAAAAGTGAACCACTCTGAGGCTGGGCC	3120	
Db	1981	GAAGTACGACAGGGCCGCTGGGAAATGTGCTGCCGTAGATGTGGAGCCCTTCGAAG	2040		3132	CAGGATGTGTCTCTCATATGAGTGTGATATTAAGACTGAGGCACAGAGGAAACG	3191	
QY	2052	CAGAGGAGAGGAGCCCTCCCTGAAGAGGGACATGCCAAGGTGCTGCAGCAGCAGC	2111		3121	CAGGATGTGTCTCTCATATGAGTGTGATATTAAGACTGAGGCACAGAGGAAACG	3180	
Db	2041	CAGAGGAGAGGAGCCCTCCCTGAAGAGGGACATGCCAAGGTGCTGCAGCAGCAGC	2100		3192	AGAACACACTGTCTCTGAGACAGTGTCTCAATGTTTCTGGCAGCTCGGCCCTCTT	3251	
QY	2112	AAGGACAGCATCCGCTTGTGTCAGCAACTGCTCCTACAGCATGAGGAGCGGACAGC	2171		3181	AGAACACACTGTCTCTGAGACAGTGTCTCAATGTTTCTGGCAGCTCGGCCCTCTT	3240	
Db	2101	AAGGACAGCATCCGCTTGTGTCAGCAACTGCTCCTACAGCATGAGGAGCGGACAGC	2160		3252	TTGATGACACTTCTCTTCCACCTTGACAGCACTGTGCCGCTCATTTCTTTAAATTTTA	3311	
QY	2172	AAGCTCAGGCACTCTTTCGAGGCTGTGGGAGGTGGTCCAGATCCGACCCCATCTTCAGC	2231					

Db 3241 TTGTATGACACTCTCTTCCACCCCTGCACAGACACATGTCGCCGTCAATCTTTTAATTTA 3300  
QY 3312 AAAGATGAATGCGCAGATCTAGTAATTCACAGATGGCTCTTGTGGGGTGGGTCTGA 3371  
Db 3301 AAAGATGAATGCGCAGATCTAGTAATTCACAGATGGCTCTTGTGGGGTGGGTCTGA 3360  
QY 3372 GGAAGTCAAGTATATAACATTTGCTGGAGTTTGTCAATGGGCTGTGCAATTTTATA 3431  
Db 3361 GGAAGTCAAGTATATAACATTTGCTGGAGTTTGTCAATGGGCTGTGCAATTTTATA 3420  
QY 3432 TTATGTTTGTAAATGACATGTCAGCGTGTGTTTCATGTTTCCATAAGCAGATATTT 3491  
Db 3421 TTATGTTTGTAAATGACATGTCAGCGTGTGTTTCATGTTTCCATAAGCAGATATTT 3480  
QY 3492 GCAACATTTGTTTGTATAGGAATTTATTTGTCACCTGCTGTGACTGTTTCTTTTGC 3551  
Db 3481 GCAACATTTGTTTGTATAGGAATTTATTTGTCACCTGCTGTGACTGTTTCTTTTGC 3540  
QY 3552 TAGTGACTAGTACCTGTTGTTCTTAAACATGAGTTTCAGCCCTTTGTTTGTAAATA 3611  
Db 3541 TAGTGACTAGTACCTGTTGTTCTTAAACATGAGTTTCAGCCCTTTGTTTGTAAATA 3600  
QY 3612 CCATGTCAAATGCAACATCTCAATCTCCCAATTTAGCTTTTAACTGACGTTCTCTTC 3671  
Db 3601 CCATGTCAAATGCAACATCTCAATCTCCCAATTTAGCTTTTAACTGACGTTCTCTTC 3660

## RESULT 3

AF387506 Homo sapiens KIAA0156 isoform mRNA 3644 bp mRNA linear PRI 24-JUN-2002  
DEFINITION Homo sapiens KIAA0156 isoform mRNA, complete cds.  
ACCESSION AF387506  
VERSION AF387506.1 GI:14579642  
KEYWORDS  
SOURCE human.

## ORGANISM

Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

## REFERENCE

1 (bases 1 to 3644)  
Liu, Y., Li, J., Kim, B.O., Pace, B.S. and He, J.J.,  
HIV-1 Tat protein-mediated transactivation of the HIV-1 long  
terminal repeat promoter is potentiated by a novel nuclear  
Tat-interacting protein of 110 kDa, Tip10  
J. Biol. Chem. 277 (26), 23854-23863 (2002)

## JOURNAL

MEDLINE  
22075130

## PUBMED

11959860

## REFERENCE

2 (bases 1 to 3644)  
Liu, Y. and He, J.J.,  
Direct Submission  
Submitted (30-MAY-2001) Department of Microbiology and Immunology,  
Indiana University School of Medicine, R4 302, 1044 W. Walnut St.,  
Indianapolis, IN 46202, USA

## JOURNAL

Location/Qualifiers

1..3644

/organism="Homo sapiens"

/db\_xref="taxon:9606"

/tissue\_type="brain"

/dev\_stage="fetal"

15..1109

/note="truncated isoform"

/codon\_start=1

/product="KIAA0156 isoform"

/protein\_id="AAK69347.1"

/db\_xref="GI:14579643"

/translation="MATAETASPEAESKAGPKADGEDEVKARRRKLSPRAVA  
AATKTMCPNDQEEGVSESDGEYAMASSAESPGYEWYDEEEKNLEIERLE  
EQLSNVDYNDCHVDLIRLLRLEKTKVRMARQMSSEIFPLTELWLEWLHLEISMA  
QDGLREHYDLFEKAVDYICPNLWIEYQYSGVIGKQKGLKVRVSFFRALSSVG  
LHMTKGLAWAYREFESAIVEAARLEKRVHSLFRQLAIPLYDEATFAEYEENSEDP  
IPESVIONKALQOLEKYKPYEALIQAEAPRLAEYQAYIDFEMKIGDPARIQLIFE  
RALVENCILVPLDIWIRYSTTESKGFICT"

## variation

1064\*1065

/replace="tcagtacctag"

BASE COUNT  
ORIGIN

Query Match 95.1%; Score 3611.2; DB 9; Length 3644;  
Best Local Similarity 99.6%; Pred. No. 0;  
Matches 3634; Conservative 0; Mismatches 3; Indels 11; Gaps 1;

QY 5 GCGTCCGATGGCGACTCGCGCGGAAACCTCGGCTTCAGAACCCGAGGCTGAGTCCCAAGC 64  
Db 8 GCGCAAGATGGCGACTCGCGCGGAAACCTCGGCTTCAGAACCCGAGGCTGAGTCCCAAGC 67  
QY 65 TGGGCCCAAGGCTGACGAGGAGGAGATGAGTTTAAAGCGCTAGGACAAGGAGAAAGT 124  
Db 68 TGGGCCCAAGGCTGACGAGGAGGAGATGAGTTTAAAGCGCTAGGACAAGGAGAAAGT 127  
QY 125 GTTATCGCGGCTGTCGCGGCTGCACATACAGACCATGGGCGGAGGCTGGATCAGCA 184  
Db 128 GTTATCGCGGCTGTCGCGGCTGCACATACAGACCATGGGCGGAGGCTGGATCAGCA 187  
QY 185 GGAGGAAGCGGTGACGAGAGCGATGGGATGAGTACGCTACGCTATGCTTCTCCGCGGAGAG 244  
Db 188 GGAGGAAGCGGTGACGAGAGCGATGGGATGAGTACGCTACGCTATGCTTCTCCGCGGAGAG 247  
QY 245 CTCGCCCGGAGTACGAGTGGGAATATACGAAGAGGAGGAGAAACCAAGCTGGAGAT 304  
Db 248 CTCGCCCGGAGTACGAGTGGGAATATACGAAGAGGAGGAGAAACCAAGCTGGAGAT 307  
QY 305 TGAGAGACTGGAGGAGGAGTGTCTATCAAGCTCTGACTACAACTGCTCATGCTGCTGACT 364  
Db 308 TGAGAGACTGGAGGAGGAGTGTCTATCAAGCTCTGACTACAACTGCTCATGCTGCTGACT 367  
QY 365 GATCAGACTGCTCAGGCTGGAGGAGGAGCTTACCAAGGTGAGGATGGCGCGGAGAGAT 424  
Db 368 GATCAGACTGCTCAGGCTGGAGGAGGAGCTTACCAAGGTGAGGATGGCGCGGAGAGAT 427  
QY 425 GAGTGAATCTTTCCTTGAAGAGCTCTGGCTGGAGTGGCTGGATGAGGAGATGAGATCAG 484  
Db 428 GAGTGAATCTTTCCTTGAAGAGCTCTGGCTGGAGTGGCTGGATGAGGAGATGAGATCAG 487  
QY 485 CATGCCCGGAGTGGCTGGAGAGACACGCTGATGACCTCTTTGAGAAGCGCTGAA 544  
Db 488 CATGCCCGGAGTGGCTGGAGAGACACGCTGATGACCTCTTTGAGAAGCGCTGAA 547  
QY 545 GGATTACATTTGCTCTAACTTTGGCTAGAGTATGGCTAGTGGCTAGTGGCTAGTGGCTAGTGG 604  
Db 548 GGATTACATTTGCTCTAACTTTGGCTAGAGTATGGCTAGTGGCTAGTGGCTAGTGGCTAGTGG 607  
QY 605 TCAGAAAGGTGGCTTGAAGAAATTCGCTCCGCTTTGAAAGGCTCTCTCGTCTGTTGG 664  
Db 608 TCAGAAAGGTGGCTTGAAGAAATTCGCTCCGCTTTGAAAGGCTCTCTCGTCTGTTGG 667  
QY 665 TTTACATATGACCAAGGACTCGCCCTCTGGGAGGCTTACCAGAGAGTTTGAAGTGCAT 724  
Db 668 TTTACATATGACCAAGGACTCGCCCTCTGGGAGGCTTACCAGAGAGTTTGAAGTGCAT 727  
QY 725 TGTGGAAGCTGCTCGGCTTGAGAAAGTCCACAGTCTTTTCGGCGGACAGTGGCGATCCC 784  
Db 728 TGTGGAAGCTGCTCGGCTTGAGAAAGTCCACAGTCTTTTCGGCGGACAGTGGCGATCCC 787  
QY 785 ACTCTATGATGAGGCGCACATTTGACAGATATCAAGATGGTTCAGAGAGCCCAATACC 844  
Db 788 ACTCTATGATGAGGCGCACATTTGACAGATATCAAGATGGTTCAGAGAGCCCAATACC 847  
QY 845 AGAGTCAGTAAATTCAGAACTATAACAAAGCACTACAGCAGCTGGAGAAATATAAACCCCTA 904  
Db 848 AGAGTCAGTAAATTCAGAACTATAACAAAGCACTACAGCAGCTGGAGAAATATAAACCCCTA 907  
QY 905 TGAAGAAGCACTGTTTGCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 964  
Db 908 TGAAGAAGCACTGTTTGCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 967  
QY 965 TTTTCAGATGAAATTTGGCGATCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1024  
|||||

Db	968	TTTTGAGATGAAATTTGGCGATCTGCTCGCATTCAGTTGATCTTTGAGCGCGCCCTGGT	1027
Qy	1025	CGAGAACTGCTTGTCCAGACTTTATGATCCGTTACAGTCAGTACCTAGATGACAACT	1084
Db	1028	CGAGAACTGCTTGTCCAGACTTTATGATCCGTTAC-----AGATCGCAACT	1076
Qy	1085	GAAAGTAAAGATTTGGTTTATCTGTACATAACCGCGCTATTAGAAACTGCCCTGGAC	1144
Db	1077	GAAAGTAAAGATTTGGTTTATCTGTACATAACCGCGCTATTAGAAACTGCCCTGGAC	1136
Qy	1145	AGTTGCTTATGGAGTCGGTACCTTTGGCCATCGAGAGACATGAGTTGATCATCAAGT	1204
Db	1137	AGTTGCTTATGGAGTCGGTACCTTTGGCCATCGAGAGACATGAGTTGATCATCAAGT	1196
Qy	1205	AATTTCTGTAACTTTCAGAAAGCTTTGAATGCCGGCTTCATCCAGGCGCACTGATTTATGT	1264
Db	1197	AATTTCTGTAACTTTCAGAAAGCTTTGAATGCCGGCTTCATCCAGGCGCACTGATTTATGT	1256
Qy	1265	GGAGATTTGGAGGCATACCTTGATTAACCTGAGGAGAAAGGTTGATTTCAAACAAGACTC	1324
Db	1257	GGAGATTTGGAGGCATACCTTGATTAACCTGAGGAGAAAGGTTGATTTCAAACAAGACTC	1316
Qy	1325	CAGTAAAGAGCTGAGAGTTGAGGCGGCTTTACTCGCTTGGAGTATCTGAAGCA	1384
Db	1317	CAGTAAAGAGCTGAGAGTTGAGGCGGCTTTACTCGCTTGGAGTATCTGAAGCA	1376
Qy	1385	GGAGTGGAGAGCGTTTCAATGAGAGTGTGTGATCCAAAGCTGCGTGATATGCAAGACTG	1444
Db	1377	GGAGTGGAGAGCGTTTCAATGAGAGTGTGTGATCCAAAGCTGCGTGATATGCAAGACTG	1436
Qy	1445	GGCTAGGATTTAGGCTCGACTGTGCAATAACATCGAGAAAGCTCGGAACTCTGGGATAG	1504
Db	1437	GGCTAGGATTTAGGCTCGACTGTGCAATAACATCGAGAAAGCTCGGAACTCTGGGATAG	1496
Qy	1505	CATCATGCCAGAGGAATGCCAAGTAGCGCAACATGTGGCTAGAGTATTACAACTGGA	1564
Db	1497	CATCATGCCAGAGGAATGCCAAGTAGCGCAACATGTGGCTAGAGTATTACAACTGGA	1556
Qy	1565	AAGAGCTCATGTTGACACCCAGCACTGCGGAAAGGCTCTGCACCGGCGCTCCAGTGCAC	1624
Db	1557	AAGAGCTCATGTTGACACCCAGCACTGCGGAAAGGCTCTGCACCGGCGCTCCAGTGCAC	1616
Qy	1625	CAGTGACTACCCAGAGCAGCTCTCGGAAGTGTACTCACCATGGAGAGGACAGAAAGTTC	1684
Db	1617	CAGTGACTACCCAGAGCAGCTCTCGGAAGTGTACTCACCATGGAGAGGACAGAAAGTTC	1676
Qy	1685	TTTAAAGATTTGGATATAGCTGTTCAAGAACTGAAACCCGATTTAGCTGTGTCAATGA	1744
Db	1677	TTTAAAGATTTGGATATAGCTGTTCAAGAACTGAAACCCGATTTAGCTGTGTCAATGA	1736
Qy	1745	GCAGAGATTAAGGCTGCAGAGAAGAGCAGCGCTTTGTGCAGCAAGAAGAAAAGGC	1804
Db	1737	GCAGAGATTAAGGCTGCAGAGAAGAGCAGCGCTTTGTGCAGCAAGAAGAAAAGGC	1796
Qy	1805	TGAACACGGAAAAAGCTCGGCTGAGAAAGAGCGTTTAAAAAGAAAGAAAAGATCAG	1864
Db	1797	TGAACACGGAAAAAGCTCGGCTGAGAAAGAGCGTTTAAAAAGAAAGAAAAGATCAG	1856
Qy	1865	AGCCCCAGAGAGCGCGAGCAGATGAGGACGATGAGAAAGAGTGGGCGCATGATGAAGA	1924
Db	1857	AGCCCCAGAGAGCGCGAGCAGATGAGGACGATGAGAAAGAGTGGGCGCATGATGAAGA	1916
Qy	1925	AGAGCAGCTTCCAAACGCAAGGGTCGAGAACAGCATCCCTGCAGCTGGAGAAACACA	1984
Db	1917	AGAGCAGCTTCCAAACGCAAGGGTCGAGAACAGCATCCCTGCAGCTGGAGAAACACA	1976
Qy	1985	AAATGTAGAGTAGCAGAGGCGCGCTGGGAAATGTGCTGCCGTAGATGTGGAGCCCC	2044
Db	1977	AAATGTAGAGTAGCAGAGGCGCGCTGGGAAATGTGCTGCCGTAGATGTGGAGCCCC	2036
Qy	2045	TTCCAGCAGAGAGAGGAGCGCTCCCTGAAGAGGACATGCCCAAGTGTGTCAGGA	2104
Db	2037	TTCCAGCAGAGAGAGGAGCGCTCCCTGAAGAGGACATGCCCAAGTGTGTCAGGA	2096
Qy	2105	CAGCAGAAAGGACAGCATCACCGTCTTTGTGAGCAAACTGCCCTACAGCATCGAGGACC	2164
Db	2097	CAGCAGAAAGGACAGCATCACCGTCTTTGTGAGCAAACTGCCCTACAGCATCGAGGACC	2156
Qy	2165	GGACACGAAGCTCAGGCCACTCTTCGAGGCTGTGGGAGGTGTGTCAGATCCAGCCCAT	2224
Db	2157	GGACACGAAGCTCAGGCCACTCTTCGAGGCTGTGGGAGGTGTGTCAGATCCAGCCCAT	2216
Qy	2225	CTTCAGCAACCGTGGGATTTCCGAGGTTACTGCTACGTGAGTTTAAAGAGAGAAATC	2284
Db	2217	CTTCAGCAACCGTGGGATTTCCGAGGTTACTGCTACGTGAGTTTAAAGAGAGAAATC	2276
Qy	2285	AGCCCTTCAGGCACCTGGAGATGGACCGGAAAGTGTAGAAGGAGGCGCAATCTTTGTTTC	2344
Db	2277	AGCCCTTCAGGCACCTGGAGATGGACCGGAAAGTGTAGAAGGAGGCGCAATCTTTGTTTC	2336
Qy	2345	CCCTGTGTGATTAAGAGCAAAAACCCCGATTTTAAAGTGTTCAGGTACAGCACTTCCTT	2404
Db	2337	CCCTGTGTGATTAAGAGCAAAAACCCCGATTTTAAAGTGTTCAGGTACAGCACTTCCTT	2396
Qy	2405	AGAGAAACACAGCTGTTTCATCTCAGGCTGCTTCTCCTCTACTAAAGAGGAACTAGA	2464
Db	2397	AGAGAAACACAGCTGTTTCATCTCAGGCTGCTTCTCCTCTACTAAAGAGGAACTAGA	2456
Qy	2465	AGAAATCTGTAAAGCTCATGTCACCGTGAAGGACCTCAGGCTGTGTCACCAACCGGCTGG	2524
Db	2457	AGAAATCTGTAAAGCTCATGTCACCGTGAAGGACCTCAGGCTGTGTCACCAACCGGCTGG	2516
Qy	2525	CAAAACAAAGGCTGCTGCTACGTGAGTATGAAAATGAATCCAGGCGTCCGAGGCTGT	2584
Db	2517	CAAAACAAAGGCTGCTGCTACGTGAGTATGAAAATGAATCCAGGCGTCCGAGGCTGT	2576
Qy	2585	GATGAAGTGCACGGCATGACTATCAAGAGAAACATCATCAAGTGGCAATCAGCAACCC	2644
Db	2577	GATGAAGTGCACGGCATGACTATCAAGAGAAACATCATCAAGTGGCAATCAGCAACCC	2636
Qy	2645	TCCTCAGAGGAAAGTTCAGAGAAAGCAGACACAGGAGGACCAAGTGGCCCATGCT	2704
Db	2637	TCCTCAGAGGAAAGTTCAGAGAAAGCAGACACAGGAGGACCAAGTGGCCCATGCT	2696
Qy	2705	TTTCCGCGAGACATACGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG	2764
Db	2697	TTTCCGCGAGACATACGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG	2756
Qy	2765	TGCCCTGCAGCGCCCAAGTGTGCTGAGTGTCTCAGGCTGAGAACGCGCTGCGCGGCTCC	2824
Db	2757	TGCCCTGCAGCGCCCAAGTGTGCTGAGTGTCTCAGGCTGAGAACGCGCTGCGCGGCTCC	2816
Qy	2825	TGCAGTTGCCCGCCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG	2884
Db	2817	TGCAGTTGCCCGCCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG	2876
Qy	2885	GCTGTTTCTGAGAAAGTGAACGGGACGCTGGGAGACAGGAAATGCTTACTCTCTG	2944
Db	2877	GCTGTTTCTGAGAAAGTGAACGGGACGCTGGGAGACAGGAAATGCTTACTCTCTG	2936
Qy	2945	CCCGGCGGACCTCCCAACCCAGCAGTGCAGTGGGATGACAGGCGTGTGCTGCTCG	3004
Db	2937	CCCGGCGGACCTCCCAACCCAGCAGTGCAGTGGGATGACAGGCGTGTGCTGCTCG	2996
Qy	3005	TGCTCGCAACCCACAGATGGCTTCTCGGCTTTAGACAGAAAGGAGGAGGAGGAGGAGG	3064
Db	2997	TGCTCGCAACCCACAGATGGCTTCTCGGCTTTAGACAGAAAGGAGGAGGAGGAGGAGG	3056
Qy	3065	AGAGCCTTTCAGTCTCCTCATATTGAGGCGAGTGGCAGAAAAGTACCACCTCTG	3124
Db	3057	AGAGCCTTTCAGTCTCCTCATATTGAGGCGAGTGGCAGAAAAGTACCACCTCTG	3116
Qy	3125	CTGGGCGGAGGATGTGCTGCTGATAGTTTGTATCTTAAAGACTGAGGCGACAGAG	3184
Db	3117	CTGGGCGGAGGATGTGCTGCTGATAGTTTGTATCTTAAAGACTGAGGCGACAGAG	3176

```
QY 3185 CGAAACGAGAACACACTGTTTTGTGACACACAGTGTGCCAAATGTTTCTGGCCAGCTCCGG 3244
Db 3177 CGAAACGAGAACACACTGTTTTGTGACACACAGTGTGCCAAATGTTTCTGGCCAGCTCCGG 3236
QY 3245 CCCCTTTTGTATGACACACTTCTTCCACCCTGCACAGCACATGTGCCCGTCAATCTTTT 3304
Db 3237 CCCCTTTTGTATGACACACTTCTTCCACCCTGCACAGCACATGTGCCCGTCAATCTTTT 3296
QY 3305 AATTTTAAAGATGAATGGCAGACTAGTATTAATTCACAGAATGGCTCTTGTGGGGTG 3364
Db 3297 AATTTTAAAGATGAATGGCAGACTAGTATTAATTCACAGAATGGCTCTTGTGGGGTG 3356
QY 3365 GGTCTGAGGAAGTCAGCTATATAACATTTGCTGGAGTTTGTTCATAGGGCTGTGCAT 3424
Db 3357 GGTCTGAGGAAGTCAGCTATATAACATTTGCTGGAGTTTGTTCATAGGGCTGTGCAT 3416
QY 3425 TTTTATATATGTTTGTAAATGACATGTCAGCCCTTGTTCATGTTTCTTAAAGCAG 3484
Db 3417 TTTTATATATGTTTGTAAATGACATGTCAGCCCTTGTTCATGTTTCTTAAAGCAG 3476
QY 3485 AATATTGCAACATTTGTTTGTATAGGAATTTATTTGCCACCTGCTGTGGACTGTTT 3544
Db 3477 AATATTGCAACATTTGTTTGTATAGGAATTTATTTGCCACCTGCTGTGGACTGTTT 3536
QY 3545 CTTTGCTAGTACAGTACCTGTTCTCTTAACATGAGTTTCAGCCCTTTGGTTTG 3604
Db 3537 CTTTGCTAGTACAGTACCTGTTCTCTTAACATGAGTTTCAGCCCTTTGGTTTG 3596
QY 3605 TTTAATACCATGTCAATCAACATTCATCTCCCATTTAGCTTTA 3652
Db 3597 TTTAATACCATGTCAATCAACATTCATCTCCCATTTAGCTTTA 3644

RESULT 4
AX268309
LOCUS AX268309 3537 bp DNA linear PAT 29-OCT-2001
DEFINITION Sequence 30 from Patent WO0174855.
ACCESSION AX268309
VERSION AX268309.1 GI:16541547
KEYWORDS synthetic construct.
SOURCE synthetic construct.
ORGANISM artificial sequences.
REFERENCE
1
AUTHORS Laus, R., Vidovic, D. and Graddis, T.
TITLE Compositions and methods for dendritic cell-based immunotherapy
JOURNAL Patent: WO 0174855-A 30 11-OCT-2001;
DENDREON CORPORATION (US)
FEATURES
source Location/Qualifiers
1..3537
/organism="synthetic construct"
/db_xref="taxon:32630"
/note="SART-3-1C"
BASE COUNT 931 a 885 c 1048 g 673 t
ORIGIN

Query Match 76.0%; Score 2886; DB 6; Length 3537;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 2886; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 12 ATGGGAGCTGGCGGCGAAACCTTCGGCTTCAGAACCCGAGGCTGAGTCCAGGCTGGGCC 71
Db 1 ATGGGAGCTGGCGGCGAAACCTTCGGCTTCAGAACCCGAGGCTGAGTCCAGGCTGGGCC 60
QY 72 AAGGCTGACGAGAGGAGGATGAGTTAAGCGGCTAGGACAGGAGGAAGGTATGCG 131
Db 61 AAGGCTGACGAGAGGAGGATGAGTTAAGCGGCTAGGACAGGAGGAAGGTATGCG 120
QY 132 CGGGCTGTGGCGCTGCGACATACAGACCATGGGGCCAGCGCTGGGATCAGCAGGAGGAA 191
Db 121 CGGGCTGTGGCGCTGCGACATACAGACCATGGGGCCAGCGCTGGGATCAGCAGGAGGAA 180
QY 192 GCGGTGAGCGAGGAGGATGAGTACGCCATGGCTTCCTCCGCGGAGAGCTCCGCC 251
```

```
Db 181 GGCGTGAGCGAGCGATGGGATGAGTACGCCATGGCTCTCTCCGCGAGAGCTCCCC 240
QY 252 GGGGAGTACAGTGGGAATATGACGAAGAGGAGGAGAAAACACAGCTGAGATTGAGAGA 311
Db 241 GGGGAGTACAGTGGGAATATGACGAAGAGGAGGAGAAAACACAGCTGAGATTGAGAGA 300
QY 312 CTGGAGGACAGTGTCTATCAACGTCTATGACTACAACTGCGCATGTGACATTGATCAGA 371
Db 301 CTGGAGGACAGTGTCTATCAACGTCTATGACTACAACTGCGCATGTGACATTGATCAGA 360
QY 372 CTGCTCAGGCTGGAAGGGAGCTTACCAAGGTGAGGATGCGCCGCGCAGAAAGATGATGAA 431
Db 361 CTGCTCAGGCTGGAAGGGAGCTTACCAAGGTGAGGATGCGCCGCGCAGAAAGATGATGAA 420
QY 432 ATCTTTCCCTTCACTCAAGAGCTCTGGCTGGAGTGGCTGCATGACAGATCAGCATGGCC 491
Db 421 ATCTTTCCCTTCACTCAAGAGCTCTGGCTGGAGTGGCTGCATGACAGATCAGCATGGCC 480
QY 492 CAGGATGGCTGACAGAGAGCAGCTGTATGACCTCTTTGAGAAAACCCGTGAAGATTAC 551
Db 481 CAGGATGGCTGACAGAGAGCAGCTGTATGACCTCTTTGAGAAAACCCGTGAAGATTAC 540
QY 552 ATTTGCTCTAACATTTGGCTAGAGTATGCCCAGTACTCAGTTGGTGGGATTTGGTCAGAA 611
Db 541 ATTTGCTCTAACATTTGGCTAGAGTATGCCCAGTACTCAGTTGGTGGGATTTGGTCAGAA 600
QY 612 GGTGGCTTTGAGAAAGTTCCGCTCGTGTGTTGAAAGGGCTCTCTCGTCTCTGTTTACAT 671
Db 601 GGTGGCTTTGAGAAAGTTCCGCTCGTGTGTTGAAAGGGCTCTCTCGTCTCTGTTTACAT 660
QY 672 ATGACCAAAGGACTCGCCCTCTGGAGGCTTACCAGAGAGTTTGAAGTCGCGATTGTGAA 731
Db 661 ATGACCAAAGGACTCGCCCTCTGGAGGCTTACCAGAGAGTTTGAAGTCGCGATTGTGAA 720
QY 732 GCTGCTCGCTTGAGAAAGTCCACAGTCTTTTCCGCGGACAGTTGGCGATCCCACTCTAT 791
Db 721 GCTGCTCGCTTGAGAAAGTCCACAGTCTTTTCCGCGGACAGTTGGCGATCCCACTCTAT 780
QY 792 GATATGGAGGCCACATTTGCAGAGTATGAAGAATGGTCAAGAACCCATATACAGAGTCA 851
Db 781 GATATGGAGGCCACATTTGCAGAGTATGAAGAATGGTCAAGAACCCATATACAGAGTCA 840
QY 852 GTAATTCAGAACTATAACAAAGCACTACAGAGCTGGAGAAATATAAACCCTTATGAAGAA 911
Db 841 GTAATTCAGAACTATAACAAAGCACTACAGAGCTGGAGAAATATAAACCCTTATGAAGAA 900
QY 912 GCACCTGTTGCAGGAGGCGCACCAAGGCTGGCAGAAATATCAAGCATATATCATTTTGAG 971
Db 901 GCACCTGTTGCAGGAGGCGCACCAAGGCTGGCAGAAATATCAAGCATATATCATTTTGAG 960
QY 972 ATGAAAATTTGGGATCCCTGCTCGCATTCAGTTTGATCTTTGAGCGGCCCTGGTCAGAAC 1031
Db 961 ATGAAAATTTGGGATCCCTGCTCGCATTCAGTTTGATCTTTGAGCGGCCCTGGTCAGAAC 1020
QY 1032 TGCCTTGTCCCAGACTTATGATCCGTTTACAGTCTAGTACTAGATCGACAACCTGAAAGTA 1091
Db 1021 TGCCTTGTCCCAGACTTATGATCCGTTTACAGTCTAGTACTAGATCGACAACCTGAAAGTA 1080
QY 1092 AAGGATTTGGTTTTTATCTGTACATACCCCGCTATTAGAACTGCCCTGGACAGTTGCC 1151
Db 1081 AAGGATTTGGTTTTTATCTGTACATACCCCGCTATTAGAACTGCCCTGGACAGTTGCC 1140
QY 1152 TTATGAGTTCGCTACCTCTTGGCCATGGAGACACATGGAGTTGATCATCAAGTAAATTTCT 1211
Db 1141 TTATGAGTTCGCTACCTCTTGGCCATGGAGACACATGGAGTTGATCATCAAGTAAATTTCT 1200
QY 1212 GTAACTTTTCGAGAAAGCTTTGAATCCGCTTTTCATCCAGGCCACTGATTATGTGGAGATT 1271
Db 1201 GTAACTTTTCGAGAAAGCTTTGAATCCGCTTTTCATCCAGGCCACTGATTATGTGGAGATT 1260
QY 1272 TGGCAGGCTATACCTTGATTACCTGAGGAGAGGGTTGATTTCAAACAGACTCCAGTAAA 1331
```



15...2903

/gene="Sart3"

/note="similar to Homo sapiens SART3"

/codon\_start=1

/product="tumor-rejection antigen SART3"

/protein\_id="AAF65228.1"

/db\_xref="GI:7637845"

/translation="MATTAASASAEPEVEPOAGPEAEGEDEEKAPAGVORVKVLSGAVA  
AABAAPKGPMDLOREGASGDDEDEAMASSAESSEDEWEDYDEEKNOLEIERL  
EBOLSINGVDYVCHVELRLRLLEGLSRVAAKQKSELFPLEELWLEWLEHETSM  
ANDGLDRHEVLEPERAVKDYICPNIMLEYGOYVSGVIGQGGLEKVRVPERALSSV  
GLHMTKGLAIMEAYPREPSAIVEAARLEKHSFRLOLAIPLEYEMFATFAEYEWSEE  
PMPESVLOSIOKALUGSKPYEALLOAPRLAETQAYIDFEMKIPDRIQIIF  
ERALVENICLFDLWIRYSOYLDRLQVKDLVLSYRAVRNCPWTVALWSKRYLLAMER  
HGLDHQTSATPENALSAGFTQAYDVEIMOVYLDLYRRVDRQDSKSELELRSMF  
TRALVELOVEERESGSDCLIMOSWARVEARLNNOKKAPRELWDSIMTRGNKY  
ANNWLEYNLERHGDTOHCKALHRAVOCTSDYPEHVCVELITWETEGTLEDWDLA  
IOKTRTLARVNEQKMAEKEALVQOEBAQRKKVRAEKALKKKKTRGADKR  
RDEDEENWGEDEEPEQSKRRRTENSLSAESAEMKEETELSKCLTIDVGPVQSK  
EKAASLKDRMPKVAHDSKSDVTFVSNLPYSIEEPEVKLRPLFEVGEVVOYRPIFS  
NRGDFRGVYIEVEPEKSAOALDELDRKIVGPRMVFVSPVDKSNPDVFRVYSYTL  
EKHLFIISGLPEFSCFTELEDICAKHGTVDLRLVTRNACKPKGLAVVEVESQASQ  
AVMKMDGHTIRENVIKVAISNPPQKVPKPEVPTAPCAPMLPQMGARGKRGKTOLS  
LIPRALQROGAPOAENPAGPAPVAPSVATEAPKMSNADFALLLRK"

BASE COUNT

846 a

901 c

1147 g

692 t

Query Match

56.7%; Score 2154.2; DB 10; Length 3586;

Best Local Similarity 77.5%; Pred. No. 0;

Matches 2807; Conservative 0; Mismatches 728; Indels 86; Gaps 13;

QY

5

GGCTCGATGGCGACTGGCGCGAACCCTCGGCTTCAGAACCAGGCGTGAAGGC 64

DB

8

GCGCAAGTGGCAGCAGCGCGCATCTTCGGCGTCGCGGAGGTGAGCCCCAGGC 67

QY

65

TGGGCCAAGGCTGACGGAGAGGAGATGAGGTTAAGCGGCTAGGACAAGGAAAGGT 124

DB

68

CGGGCTTGAGCGGAGGAGAGAGATGAGCGAAGCGCGCGGTGTGCAGCGGAAGGT 127

QY

125

GTATTCGGCGGCTGTGGCGGTGGCACATACAAGACCATGGGCGACGCTGGGATCAGCA 184

DB

128

GCTGTCGGCGCTGTGGCGGCGGAGCGGAGGCCAAGGGCCCCGGATGGGACCTGCA 187

QY

185

GGAGGAAGCGCTGACGAGAGCGATGGGGAT---GAGTACGCCATGGCTTCCTCCGCGGA 241

DB

188

GCGGGAAGCGCAGCGGAGCGATGGGGATGAGGAGGACGCCATGCTTCCTCCGCGGA 247

QY

242

GAGCTCCCCGGGAGTACGAGTGGGAATATGACAAGAGGAGGAGAAAACCAAGCTGGA 301

DB

248

GAGCTCCGCGGGAGGACGAGTGGGAGTACGAGGAGGAGGAGGAAGAACCAAGCTGGA 307

QY

302

GATTGAGACTGGAGGAGCAGTTGCTCTATCAACGCTATGACTACAACTGCCATGTGA 361

DB

308

GATCGAGCGGCTGGAGGAGCAGCTGTTCATCAATGGCTATGATTACAACTGCCACGTGA 367

QY

362

CTTGATCAGACTGCTCAGCTGGAGGAGGAGCTTACCAAGGTAGGATGGCCGCCAGAA 421

DB

368

GCTCATCAGGCTGCTCGGCTGGAGGCGAGCTCAGCAGAGTGAAGGCGGCCGCCAGAA 427

QY

422

GATGAGTGAATCTTTCCCTTGACTGAAGAGCTCTGGCTGGAGTGGCTGCATGACGAGAT 481

DB

428

GATGAGTGAAGCTCTTCCCCCTGACCGAAGAGCTCTGGCTGGAGTGGCTCCACGATGAGAT 487

QY

482

CAGCATGCCCAGGATGGCTGGACAGAGACAGTGTATGACCTCTTTGAGAAAGCCGT 541

DB

488

CAGCATGGCCATGGAGGCGCTGGACCGGAGCAGCTGTACGAGCTCTTTGAGAGAGCCGT 547

QY

542

GAAGGATTACATTTGCTCCACATTTGGCTAGAGTATGGCAGTACCTACGTTGGTGGGAT 601

DB

548

GAAGGACTACATCTGTCACACATTTGGCTAGAGATATGGCCAGTACTACGTTGGTGGCAT 607

QY

602

TGGTCAGAAAGGTGGCCCTTGAGAAAGTTTCGCTCCGCTGTTTGAAGGGCTCTCTCGTCTGT 661

608

TGGTCAGAAAGGTGGCCCTTGAGAAAGGTTGCGTCTGCTCTGTTTGAAGAGCCCTGCTCTCTGT 667

QY

662

TGGTTTACATATGACCAAAAGGACTCGCCCTCTCGGAGGCTTACCGAGAGTTTGAAGTGC 721

DB

668

TGGCCTGCACATGACGAAAGCCCTGGCCATCTGGGAGGCCCTACCGAGAGTTTGAAGCGC 727

QY

722

GATTTGGAAGCTGCTCGGCTTGAGAAAGTCCACAGTCTTTTCCCGGCGACAGTTGGCGAT 781

DB

728

CATCGTGGAGGCTGCTCGGCTGGAGAAAGTGCACAGTCTCTCTCGGCGACAGCTGGCGAT 787

QY

782

CCCACTCTATCATATGGAGGCCACATTTTGAGAGATATGAAGATGTCAGAGAACCAAT 841

DB

788

CCCAGTACAGAGATGGAGGCCACCTTTGACAGATATGAAGATGTCAGAGAGGCCAT 847

QY

842

ACCAGAGTCAGTAATTCAGAACTATAACAAGACACTACACAGCTGGGAGAAATATAAAC 901

DB

848

GCGGAGTCTGTACTTCAGACTATCAGAAAGCCCTGGGCGAGCTAGAGAGTACAGGCC 907

QY

902

CTATGAAGAAGCACTGTTTGCAGGCGAGGACCACCAAGCTGGCGAATATCAAGCATATAT 961

DB

908

TTACGAGCAAGCGTGTCTGACAGGAGGCCCTCGCTGGCGAATACCAAGCTTACAT 967

QY

962

CGATTTTCAGATGAAATTTGGCGATCTGCTGCGATTCAGTTGATCTTTGAGCGCGCCT 1021

DB

968

CGACTTCGAGATGAAATTCGGGGATCTCGCCGCTATTTCAGTTGATCTTTGAGCGTCCCT 1027

QY

1022

GGTCGAGAACTCGCTTGCACAGACTTATGGATCCGTTACAGTCAGTACCTACATCGACA 1081

DB

1028

GGTGAGAACTCGCTTGCAGACTTATGGATCCGCTACAGTCAGTACCTACATCGACA 1087

QY

1082

ACTGAAAGTAAAGGATTTGGTTTATCTGTACATAACCGCGCTATTAGAAACTGCCCCCTG 1141

DB

1088

GCTGAAAGTAAAGGATTTGGTTTATCTGTACATAACCGCGTGTGTCGGAAGTGGCCATG 1147

QY

1142

SACAGTTGCCCTATTGGAGTGGTACCTCTTGGCCATGGAGACATGGAGTTGATCATCA 1201

DB

1148

GACAGTTGCCCTATTGGAGTGGTACCTCTTGGCCATGGAGACATGGAGTTGATCATCA 1207

QY

1202

AGTAATTTCTTAACCTTCGAGAAAGCTTTGAAATGCGCGCTTCAATCGAGTACCTACGAGCCACTGATTA 1261

DB

1208

AACGATTTCTCGGACCTTCGAGAAAGCTCTGAGTGGCGGCTTCAATCGAGTACCTACGAGCCACTGACATA 1267

QY

1262

TGTGAGATTTGGCAGGCGATACCTTGTATACCTGAGGAAAGGTTGATTTCAAAAGAA 1321

DB

1268

TGTGAGATCTGGCAGGCTGTACCTGACCTTACCTGAGGAAAGGTTGACTTCAGACAGGA 1327

QY

1322

CTCAGATTAAGAGCTGGAGGATTTGAGGCGCGCTTTTACCTGCTGCTGAGTATCTGAA 1381

DB

1328

CTCTAGCAAGAGCTGGAAAGCTGCGGTCATGTTTCACTGAGTACCTGAGTACCTGCA 1387

QY

1382

GCAGAGTGGAAAGAGCTTTCAATGAGTGGTGTATCCAAAGCTGCTGATTTATGAGAA 1441

DB

1388

GCAGAGTGGAAAGAGCTTTCAATGAGTGGTGTATCCAAAGCTGCTGATTTATGAGAA 1447

QY

1442

CTGGCTAGGATTTGAGGCGCTGCAATTAACATGAGGAAAGCTGCGGAACTCTGGGA 1501

DB

1448

CTGGCTCGGCTGAGGCTCGCCTGTGCATATACATGCAAGAACCCGAGAGCTCTGGGA 1507

QY

1502

TAGCATATGACAGAGAAATGCCAAGTACGCCAATATGAGTGGCTAGAGTATTTACAACT 1561

DB

1508

CAGCATATGACAGAGGAAATGCCAAGTACGCCAATATGAGTGGCTAGAGTATTTACAACT 1567

QY

1562

GGAAGAGCTCATGCTGACAGGCTGCGGAGGCTTCCAGGAGGCTTCCAGTGG 1621

DB

1568

GGAACGGGACACGGTGACACAACTGTGCGAAGGCTTCTGCACCGGAGCTGTCAGTG 1627

QY

1622

CACAGTACTTACCCAGGAGCGCTCTCGGAGGCTTACTCACCATGGAGGAGCAGAGG 1681

DB

1628

CACGAGTACTACCCAGGAGCGCTCTGTGAAGTGTGCTCTACCATGGAGGAGCAGAGG 1687

QY

1682

TTCCTTTAGAGATTTGGGATATGAGCTGTTTGAAGAACTGAAACCCGATTTAGTCTGCTGCA 1741

DB

1688

GACCTTTAGAGATTTGGGATATGAGCTGTTTGAAGAAAGGAGACGCGCTTGGCTGCTGAA 1747





David, R., Delgado, O., Deshazo, D., Ding, Y., Domah-Rashid, N., Dugan-Rocha, S., Durbin, K.J., Fernandez, C., Ferraguto, D., Forcum-Tansey, J., Frantz, P., Ganesh, R., Gorrell, J.H., Gorrell, L.L., Guevara, W., Harris, K., Hernandez, J., Hodgson, A., Hoques, M., Holloway, C., Hosak, H., Issar, A., Jackson, L.E., Jackson, L., Jia, Y., Jones, M., Kelly, S., Kneitz, S., Kondejewski, N., Kong, Y., Kovar, C., Lau, S., Leal, B., Lee, E., Li, Z., Lichtarge, O., Liu, J., Liu, W., Logan, O., Lozano, R.J., Lu, J., Lucier, R., Marondel, I., Martin, R., Martinez, C., McLeod, M.P., Mei, G., Merscher, S., Miller, A., Montgomery, K.T., Morgan, M., Morris, S., Nash, S., Nelson, A., Nguyen, R., Nguyen, N., Nguyen, S., Oswal, G., Parish, B., Paxton, S., Payton, B., Perez, L., Pu, L.L., Quiles, M., Reiter, D., Rives, M., Samuel, S., Say, J., Scherer, S., Shah, E., Shen, H., Shim, C., Simon, M., Sparks, A., Stamps, A., Sugang, R., Tabor, P., Taylor, T., Vasquez, L., Vinson, R., Vo, Q., Wabnah, M., Watlington, S., Weinstein, G., Weinstein, I.R., Williamson, A., Worley, K., Wren, J., Wrensford, G., Xiang, A.M., Yang, R., Yu, W., Zhou, X., Kucheriapatti, R., Nelson, D. and Gibbs, R.

## TITLE

## JOURNAL

## REFERENCE

## AUTHORS

## TITLE

## JOURNAL

## REFERENCE

## AUTHORS

## TITLE

## JOURNAL

## REFERENCE

## AUTHORS

## TITLE

## JOURNAL

## COMMENT

## Direct Submission

## Unpublished

2 (bases 1 to 171523)

## Worley, K.C.

## Direct Submission

Submitted (15-SEP-1999) Human Genome Sequencing Center, Department

of Molecular and Human Genetics, Baylor College of Medicine, One

Baylor Plaza, Houston, TX 77030, USA

3 (bases 1 to 171523)

## Worley, K.C.

## Direct Submission

Submitted (08-DEC-1999) Human Genome Sequencing Center, Department

of Molecular and Human Genetics, Baylor College of Medicine, One

Baylor Plaza, Houston, TX 77030, USA

4 (bases 1 to 171523)

## Worley, K.C.

## Direct Submission

Submitted (31-MAR-2001) Human Genome Sequencing Center, Department

of Molecular and Human Genetics, Baylor College of Medicine, One

Baylor Plaza, Houston, TX 77030, USA

On Dec 8, 1999 this sequence version replaced gi:6513881.

INFORMATION: <http://www.hgsc.bcm.tmc.edu/> or email

gc-help@bcm.tmc.edu

CLONE LENGTH: This sequence does not necessarily represent the entire insert of this clone. Overlapping regions of clones are only sequenced and submitted once, so the sequence for the remainder of the insert may be found in the record for the adjacent clones. Overlapping clones are noted at the beginning and end of the Features listing.

## ANNOTATION OF FEATURES:

STSS are identified using ePCR (Genome Res. 7:541-550) searches of a local database that includes entries from dbSTS, GDB, and local mapping efforts.

Repeats are identified using RepeatMasker (A. Smit and P. Green, unpublished.) for Human and Mouse sequences.

Genes and Region of sequence similarity are identified by BLAST (Nuc. Acids Res. 25:3389-3402) similarity (expect < 1e-34) to the EST and cDNA sequences. Genes demonstrate at least two exons flanked by consensus splice sites that maintained sequence continuity across the splice junctions. Sequences that are not identical matches are annotated as similar.

SEQUENCING READ COVERAGE: Sequencing is completed to a minimum standard of double strand coverage with a minimum of 2 clones and 2 reads with no ambiguities or 2 chemistries with a minimum of 2 clones and 3 reads with no ambiguities. If the sequence quality for a region does not meet this standard, it will be indicated in the annotation as Low Coverage.

QUALITY OF INDIVIDUAL BASES: This sequence meets stringent quality standards - estimated error rate less than 1 per 10,000 bases. Reports of lowest quality individual bases and measures of base quality are listed below. Description of the metrics can be found

at URL: <http://gc.bcm.tmc.edu:8088/quality.info/genbank.annotation.html>.

## QUALSTAT-REPORT-----

```
----- Summary Statistics -----
Contig length: 171523
Phrap values in estimate: 170934
Average error rate (BCM-Phrap estimate): 0.000527782
Fraction of Phrap values less than 40 : 0.0648358
Number of consensus changing edits: 29
Number of N's in consensus : 0
```

```
----- Consensus changing edits -----
Position      Original+Context      Edited+Context
1245          tgtaactatt(n)gtatatactg      tgtaactatt(t)gtatatactg
17241         attgtttttt(n)tttgtttgt      attgtttttt(t)tttgtttgt
17242         ttgtttttt(n)tttgtttgt      ttgtttttt(g)tttgtttgt
19260         caaaaaaaa(n)aaaaaaga      caaaaaaaa(a)aaaaaaga
19262         aaaaaaaa(n)aaaaaaga      aaaaaaaa(a)aaaaaaga
32246         cctaggtga(n)ggcgtgtgcc      cctaggtga(g)ggcgtgtgcc
52591         gagtgggtgt(n)tgattaggt      gagtgggtgt(c)tgattaggt
67506         ctccctttc(a)taccctccc      ctccctttc(c)taccctccc
84775         atttttga(n)tttttttt      atttttga(a)tttttttt
84913         gcgtgagcca(n)tgtcttggc      gcgtgagcca(c)tgtcttggc
84958         tctctccca(a)agtgctcac      tctctccca(g)agtgctcac
85951         tcttttttt(n)tttttttt      tcttttttt(t)tttttttt
85958         aaaggatgc(n)taagtnaaga      aaaggatgc(t)taagtnaaga
89544         atgntaagt(n)aaatannng      atgntaagt(t)aaatannng
89551         agtnaagata(n)nggggttgt      agtnaagata(t)nggggttgt
89552         gtnaagatan(n)gggggttgt      gtnaagatan(g)gggggttgt
89638         aaatactct(n)tttgacttt      aaatactct(t)tttgacttt
113724        ttctctct(n)ttctctct      ttctctct(t)ttctctct
113738        ttctttct(n)ttctttct      ttctttct(t)ttctttct
127359        gggtttattg(n)tttttttt      gggtttattg(t)tttttttt
127561        tttaattgtt(n)tttttttt      tttaattgtt(t)tttttttt
127590        tttttttt(a)aaagctgcc      tttttttt(t)aaagctgcc
162732        ccattggaa(n)gactccaaag      ccattggaa(c)gactccaaag
162765        atggccggag(n)tgccntctg      atggccggag(c)tgccntctg
162771        ggagntgccc(n)tcgtccagtt      ggagntgccc(c)tcgtccagtt
162860        tcaatctgtt(n)nancaccaac      tcaatctgtt(g)caacccaacc
162861        caatctgtt(n)ancccaacc      caatctgtt(g)aaacccaacc
162863        atctgttga(n)cccaaacctc      atctgttga(c)cccaaacctc
```

## ----- Distribution of Quality &lt; 40 Bases -----

```
1000|      *      *      *      *      *
900|      *      *      *      *      *
800|      *      *      *      *      *
700|      *      *      *      *      *
600|      *      *      *      *      *
500|      *      *      *      *      *
400|      *      *      *      *      *
300|      *      *      *      *      *
200|      *      *      *      *      *
100|      *      *      *      *      *
0|      *      *      *      *      *
```

5 10 15 20 25 30 35 40  
Phrap Value Range

Version: 1.01 qkfo

----- Summary Statistics -----



Scherer, S., Shah, E., Shen, H., Shim, C., Simon, M., Sparks, A., Stamps, A., Sugang, R., Tabor, P., Taylor, T., Vasquez, L., Vinson, R., Vo, Q., Wahbah, M., Watlington, S., Weinstock, G., Weinstock, I. R., Williamson, A., Worley, K., Wren, J., Wrenford, G., Xiang, A. M., Yang, R., Yu, W., Zhou, X., Kucherlapati, R., Nelson, D. and Gibbs, R.

Direct Submission  
Unpublished  
2 (bases 1 to 173153)  
Worley, K.C.

Direct Submission  
Submitted (24-JUL-1999) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA  
3 (bases 1 to 173153)  
Worley, K.C.

Direct Submission  
Submitted (28-SEP-1999) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA  
4 (bases 1 to 173153)  
Worley, K.C.

Direct Submission  
Submitted (09-OCT-1999) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA  
5 (bases 1 to 173153)  
Worley, K.C.

Direct Submission  
Submitted (16-MAY-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

On Sep 28, 1999 this sequence version replaced gi:5776564.  
INFORMATION: <http://www.hgsc.bcm.tmc.edu/> or email [gc-help@bcm.tmc.edu](mailto:gc-help@bcm.tmc.edu)

CLONE LENGTH: This sequence does not necessarily represent the entire insert of this clone. Overlapping regions of clones are only sequenced and submitted once, so the sequence for the remainder of the insert may be found in the record for the adjacent clones. Overlapping clones are noted at the beginning and end of the Features listing.

#### ANNOTATION OF FEATURES:

STSs are identified using ePCR (Genome Res. 7:541-550) searches of a local database that includes entries from dbSTS, GDB, and local mapping efforts.

Repeats are identified using RepeatMasker (A. Smit and P. Green, unpublished.) for Human and Mouse sequences.

Genes and Region of sequence similarity are identified by BLAST (Nuc. Acids Res. 25:3389-3402) similarity (expect < 1e-34) to the EST and cDNA sequences. Genes demonstrate at least two exons flanked by consensus splice sites that maintained sequence continuity across the splice junctions. Sequences that are not identical matches are annotated as similar.

SEQUENCING READ COVERAGE: Sequencing is completed to a minimum standard of double strand coverage with a minimum of 2 clones and 2 reads with no ambiguities or 2 chemistries with a minimum of 2 clones and 3 reads with no ambiguities. If the sequence quality for a region does not meet this standard, it will be indicated in the annotation as Low Coverage.

QUALITY OF INDIVIDUAL BASES: This sequence meets stringent quality standards - estimated error rate less than 1 per 10,000 bases. Reports of lowest quality individual bases and measures of base quality are listed below. Description of the metrics can be found at URL: <http://gc.bcm.tmc.edu:8088/quality.info/genbank.annotation.html>.

QUALSTAT-REPORT-----

----- Summary Statistics -----

Contig length: 173153  
Phrap values in estimate: 172401  
Average error rate (BCM-Phrap estimate): 0.00022158  
Fraction of Phrap values less than 40: 0.0281147  
Number of consensus changing edits: 43  
Number of N's in consensus: 0

Position	Original+Context	Consensus changing edits	Edited+Context
8572	caaaaaaaa(n)aaaaatagcc		caaaaaaaa(a)aaaaatagcc
23303	ggggttgagg(n)gggtcaggtt		ggggttgagg(g)gggtcaggtt
35436	aactggctat(n)acatcagaaa		aactggctat(t)acatcagaaa
59324	ttctttttt(n)ttttttttt		ttctttttt(t)ttttttttt
59488	aggctaatt(n)ntgtattttt		aggctaatt(t)ntgtattttt
59489	ggctaatt(n)tgatttttta		ggctaatt(t)tgatttttta
78531	caacccaggtt(n)cacaccagg		caacccaggtt(c)cacaccagg
78562	cacccaggg(n)tcattcaggt		cacccaggg(c)tcattcaggt
78594	ctgattatga(n)gtctctgtta		ctgattatga(t)gtctctgtta
78612	ttacattagc(n)attctctcac		ttacattagc(a)attctctcac
78637	atcattactc(n)ccaaaaggaa		atcattact(c)ccaaaaggaa
78717	ttggaggcc(n)agatgggtgg		ttggaggcc(c)agatgggtgg
78731	tggttgatc(n)cnaggtcagg		tggttgatc(a)cnaggtcagg
78733	ggttgatcnc(n)aggtcagagg		ggttgatcnc(g)aggtcagagg
84931	ctggctttgt(n)ttatntttta		ctggctttgt(t)ttatntttta
84936	ttgtnttat(n)ttttatccta		ttgtnttat(t)ttttatccta
88116	taagaatctg(n)ctttcacaat		taagaatctg(a)ctttcacaat
89153	gggaagaat(n)ttaatctaatt		gggaagaat(t)ttaatctaatt
89301	ggagtatgag(n)ccagctcggc		ggagtatgag(a)ccagctcggc
89333	aaccccgctc(n)inctaaaaat		aaccccgctc(c)inctaaaaat
89335	cccgctctnt(n)ctaaaaatac		cccgctctc(t)ctaaaaatac
89474	tactccagc(n)tggttgaaag		tactccagc(g)tggttgaaag
90120	ttttttttt(g)ggaaaaaaca		ttttttttt(t)ggtagagaca
90123	ttttttggg(a)aaaaacaggg		ttttttggg(t)aaaaacaggg
90125	tttttgagg(a)aaacagggtt		tttttgagg(t)aaacagggtt
90127	tttggaataa(a)acagggttgt		tttggaataa(g)acagggttgt
106262	tctcaaaaa(n)aaaattaata		tctcaaaaa(a)aaaattaata
110744	tccgcttga(n)gcggcgccgc		tccgcttga(g)gcggcgccgc
110782	gagcccccgc(n)tgcccgccgc		gagcccccgc(c)tgcccgccgc
110866	tcgcgactca(n)cgaggcctg		tcgcgactca(g)cgaggcctg
112388	atcattatga(n)aatctagaaa		atcattatga(a)aatctagaaa
126726	ccatcacac(n)acnaccatca		ccatcacac(c)acnaccatca
126729	tcacacnac(n)accatcacaa		tcacacnac(c)accatcacaa
126741	ccatcacac(a)accacaccca		ccatcacac(c)accacaccca
126744	tcacacac(n)accacatca		tcacacac(c)accacatca
126783	ccacacac(n)ttctccatct		ccacacac(c)ttctccatct
126846	ctttacacat(n)accacacct		ctttacacat(c)accacacct
126963	tttccactgt(n)accactcca		tttccactgt(c)accactcca
145620	cacaccgact(n)acacagcact		cacaccgact(c)acacagcact
145676	caatacacag(n)gaacacacac		caatacacag(a)gaacacacac
146199	tctcccgctc(a)ctctcccgct		tctcccgct(c)ctctcccgct
146210	ctctcccgct(a)ctctccctct		ctctcccgct(c)ctctccctct
155221	ctcccatct(n)ccccctcct		ctcccatct(c)ccccctcct

----- Distribution of Quality < 40 Bases -----

# bases	5	10	15	20	25	30	35	40
1000	*	*	*	*	*	*	*	*
900	*	*	*	*	*	*	*	*
800	*	*	*	*	*	*	*	*
700	*	*	*	*	*	*	*	*
600	*	*	*	*	*	*	*	*
500	*	*	*	*	*	*	*	*
400	*	*	*	*	*	*	*	*
300	*	*	*	*	*	*	*	*
200	*	*	*	*	*	*	*	*
100	*	*	*	*	*	*	*	*
0	*	*	*	*	*	*	*	*

Phrap value Range

[illegible]





Db 61382 TGAAGTGCCTACATAGAGCTGCTTAGCTTGAGAAATCTATATCTATGCTCCAGTGACA 61323  
Qy 773 GTTGGCAGTCCACTCATATGATGAGGCGCACATTTGCGAGAGTATGAAGATGGTTCAGA 832  
Db 61322 TTTGGTATACCACTCATGAATG-----AAATTTGCATATGTTGGAAGACTGGTTCAGA 61269  
Qy 833 AGACCCAAATACAGAGTCAAGTAATTCAGAACTATAACAAAGCACTACAGAGCTGGAGAA 892  
Db 61268 AGACCTAAGAGCAGCATTCAGCAATCCAGAACTATATAA-----CCCTGCGAGAGTTGAA 61214  
Qy 893 ATATAACCCATGAAGAGCAGTGTGCGAGGCGAGAGCCAAAGGCTGGCAGAAATATCA 952  
Db 61213 AAAAAAATACCACTAAGACATAGTTTTCAGAGCTGAGGTTCTGAGGTTGGCAGAAATGCCA 61154  
Qy 953 AGCATATATGATTTTGGAGATGAAATGGCGATCTGCTGCGCATTCAGTTGATCTTTGA 1012  
Db 61153 ACTACATAT---ATTTTGAAGCAATGGATATATCTGCTTGCATTCAGTTGATCTTTGA 61097  
Qy 1013 GCGCGCCCTGGTCGAGAACTGCCCTTGTCCAGAGCTTATGGATCGGTTACAGTCACTACCT 1072  
Db 61096 GTATGGCTGGTTAAGAAATGCCCTTGGCCACACTTAGAGA---CTACTGACAGTACCT 61041  
Qy 1073 AGATCGACACTGAAGTAAGGATTTGGTTTATCTGTATCA---TAACCGCGCTATATG 1129  
Db 61040 AGATGGAAGCTGAAAGTAATATCTCGGTTTCTTATTCACAGCAATGCCACTGGTAG 60981  
Qy 1130 AAATGCGCCCTGGAGAGTGGCTTATGGAGTCGTACTCTTGGCCACTGGAGAGACATGG 1189  
Db 60980 AAATGTTCTGGGACAAAATATTTGTAGATTTGTATCTCTTGACATCTGTAACAAATGA 60921  
Qy 1190 AGTTGATCATCAAGTAATTTCTGTAACTTTCGAGAAAGCTTTTGAATCCGCGCTTCATCCA 1249  
Db 60920 AGCTGATCATCAGATGACTTTTGTGGTCTTTGGGAAAGCTCTGATGCTGGCTTCATCTA 60861  
Qy 1250 GGCACACTGATATGTGAGATTTGGCAGGCACTACTTGATTTACCTGAGGAGAAAGGTTGA 1309  
Db 60860 GGCCATGTATAT---GGGATTTGGCTGGTTTACTTGTATGTTGCCT-----AAGGGCTG 60811  
Qy 1310 TTTCAACAAGACTCCAGTAAAGAGCTGGAGAGTGTAGGCGCCCTTTACTCTGCTGCTT 1369  
Db 60810 ATTCAACAGACTATATTAAGAG--AGTTTGGTTATTTATTCACAGTCTTTTGGGG 60753  
Qy 1370 GGAGTATCTGAAGCAGAGGTGGAAGAGCGTTTCAATGAGAGTGGTGTATCCAAAGCTGCGT 1429  
Db 60752 TATCCATCTGCTTCAGAAAGTTGAAGGATATTTCAAT---AGTGGAGTTTCAAGGCTCTT 60696  
Qy 1430 GATTATGCAGAACTGGCTAGGATTTAGGCTGCGACTGTGCAATACATGCAAGAGCTCG 1489  
Db 60695 GGTAAAGCAGAAATTTGGATCAGGACTGAGGCTCACTGCACCTTAATGTGAAGAACCTCA 60636  
Qy 1490 GGAACCTCTGGATAGCATATGACAGAGGAAATGCCAAGTACGCCAATGTGGCTAGA 1549  
Db 60635 GAAACGAGGACAGACATACACCAAGACCAAG-----AATGTATAGCTTCA 60587  
Qy 1550 GTATTAACTGAAAGAGCTCATGTGTACACCCAGCACTGCCGGAAGGCTGTGCAACCG 1609  
Db 60586 TTGTTATACACT-GAAAGAAATGCTCAGTGTGACACAGAATTTGCTAG-AGGTTCTACACTG 60529  
Qy 1610 GGCCTGTCAGTGCACAGTACTACCCAGAGCAGTGTGCGAAGTGTGTACTCACCATGGA 1669  
Db 60528 GGTGGAG-----AGGACTCATCCCGACATGTCTGTGAATCTGTGTACTCAATATGGA 60478  
Qy 1670 GAGCAGACAGGTTCTTTTAGAAGATTTGGATATAGCTGTCTCAGAAACTGAACCCGAT 1729  
Db 60477 GAGATAG-AGTCAATGAAGAGCTGATGATGTGGTGTCTGAAACTGAAC-----TC 60423  
Qy 1730 AGCTGCTGCTCAATGAGCAGAAATGAAG 1758  
Db 60422 AACTCATGCTCATGATGTTGAATGAAGG 60394

RESULT 10

AC126267/c  
LOCUS  
DEFINITION  
ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
AUTHORS  
TITLE  
JOURNAL  
COMMENT

AC126267  
Mus musculus chromosome UNK clone R23-191M12, WORKING DRAFT  
SEQUENCE, 14 unordered pieces.  
AC126267.1 GI:21693979  
HTG; HTGS\_PHASE1; HTGS\_DRAFT; HTGS\_FULLTOP.  
house mouse.  
Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
1 (bases 1 to 319737)  
McPherson,J.D. and Waterston,R.H.  
The sequence of Mus musculus clone  
Unpublished  
2 (bases 1 to 319737)  
McPherson,J.D. and Waterston,R.H.  
Direct Submission  
Submitted (04-JUL-2002) Genome Sequencing Center, 4444 Forest Park  
Parkway, St. Louis, MO 63108, USA

----- Genome Center -----  
Center: Washington University Genome Sequencing Center  
Center code: WUGSC  
Web site: http://genome.wustl.edu/gsc/index.shtml  
Contact: submissions@watson.wustl.edu  
----- Project Information -----  
Center project name: M\_BA0191M12  
----- Summary Statistics -----  
Sequencing vector: M13; 0%  
Chemistry: Dye-primer ET; 0% of reads  
Chemistry: Dye-terminator Big Dye; 100% of reads  
Assembly program: Phrap; version 0.990319  
Consensus quality: 309491 bases at least Q40  
Consensus quality: 311769 bases at least Q30  
Consensus quality: 313307 bases at least Q20  
-----  
\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 14 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.

1 1316: contig of 1316 bp in length  
\* 1317 1416: gap of unknown length  
\* 1417 3096: contig of 1680 bp in length  
\* 3097 3196: gap of unknown length  
\* 3197 8084: contig of 4888 bp in length  
\* 8085 8184: gap of unknown length  
\* 8185 14192: contig of 6008 bp in length  
\* 14193 14292: gap of unknown length  
\* 14293 18775: contig of 4483 bp in length  
\* 18776 18875: gap of unknown length  
\* 18876 27060: contig of 8185 bp in length  
\* 27061 27160: gap of unknown length  
\* 27161 36165: contig of 9005 bp in length  
\* 36166 36266: gap of unknown length  
\* 36267 42770: contig of 6505 bp in length  
\* 42771 42870: gap of unknown length  
\* 42871 83818: contig of 40948 bp in length  
\* 83819 109449: contig of 25531 bp in length  
\* 109450 109549: gap of unknown length  
\* 109550 176791: contig of 67242 bp in length  
\* 176792 176891: gap of unknown length  
\* 176892 219804: contig of 42913 bp in length  
\* 219805 219904: gap of unknown length  
\* 219905 318952: contig of 99048 bp in length  
\* 318953 319052: gap of unknown length

```
FEATURES * 319053 319737: contig of 685 bp in length.
source Location/Qualifiers
1. .319737
/organism="Mus musculus"
/db_xref="taxon:10090"
/chromosomes="UNK"
/clone="RP23-191M12"
1. .1316
/misc_feature /note="assembly_name:Contig13"
1417. .3096
/misc_feature /note="assembly_name:Contig14"
3197. .8084
/misc_feature /note="assembly_name:Contig15"
8185. .14192
/misc_feature /note="assembly_name:Contig16"
14293. .18775
/misc_feature /note="assembly_name:Contig17"
18876. .27060
/misc_feature /note="assembly_name:Contig18"
27161. .36165
/misc_feature /note="assembly_name:Contig19"
36266. .42770
/misc_feature /note="assembly_name:Contig20"
42871. .83818
/misc_feature /note="assembly_name:Contig21"
83919. .109449
/misc_feature /note="assembly_name:Contig22"
109550. .176791
/misc_feature /note="assembly_name:Contig23"
176892. .219804
/misc_feature /note="assembly_name:Contig24"
219905. .318952
/misc_feature /note="assembly_name:Contig25"
319053. 319737
/misc_feature /note="assembly_name:Contig8"
BASE COUNT 91273 a 65131 c 64726 g 97306 t 1301 others
ORIGIN
Query Match 8.9%; Score 338.2; DB 2; Length 319737;
Best Local Similarity 59.5%; Pred. No. 8.7e-71;
Matches 1088; Conservative 0; Mismatches 578; Indels 163; Gaps 24;
QY 17 GACTCGGCGGAAACCTCGGCTTCAGAACCCGAGCTGAGTCCAGGCTGGGCCCAAGGC 76
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
199732 GACAGCAGGTGCAACCCCACTTTGGAGCCAGACTAGTCTGCTGGGTACAAAGA 199673
QY 77 TGACGAGAGGAGGATGAGTTAAGCGCGCTAGGACAGAGAGAAAGTGTATCGCGGC 136
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
199672 GGAAAAGAGGAGATGAATTGAATGACGCTGGATGAGAGGATGTGACATCATGGGC 199613
QY 137 TGTGGCGCTGCGCACATACAAGACCATGGGCGAGCGTGGATCAGCAGGAGGAAGCGT 196
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
199612 TGTGGCATGAGG---TACAAGAGCATGGGCCAGATGGATCAGCAG--TGAAGGCAC 199558
QY 197 GAGCGAGCAGATGGGATGAGTAC---GCCATGGCTTCCCTCCGCGGAGAGCTCCCGCG 253
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
199557 AAGTGAGAGAATGGGATGAGGACTGTGTCTTGTGACTTCTTGGGAGGACTCTCTTGG 199498
QY 254 GGATGACGATGGGAATATGACNAGAGGAGGAGGAACCAAGC----- 297
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
199497 GAATGCAAGGAATGAGAGGATCTGGAGTAGCCAAAGAAATATCTATAAGATCTAT 199438
QY 298 -----TGGAGATTGAGAGACTGGAGGAGCAGTGTCTAT 331
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
199437 GATAACCTAAACCAATGCACCATTTTATAAATTCACACGTTGAAAAGAAATGGAATA 199378
QY 332 CAAGCTCTATGACTACAACTGCCATGTCGACTTGATCAGACTGCTCAGGCTGGAAGGGA 391
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
199377 GAAATCCCAATAAATCTTTGTGTGTGTGACCTGAACAGGCTGCTGCAGTTAGAATGAGA 199318
QY 392 GCTTACCAAGGTGAG-GATGGCCCGCCAGAACATGAGTGAATCTTCCCTTGACTGAG 450
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
199317 GCTGGCTATGGTAAGTGATAGAATCTCTATAAGAGAGACAGTCTTCCCTCTTCCCTCT 199262
```







REFERENCE	CB10 1RQ, England. E-mail contact: humquery@sanger.ac.uk
AUTHORS	2 (bases 1 to 218)
TITLE	Cross, S.H., Charlton, J.A., Nan, X. and Bird, A.P.
JOURNAL	Purification of CpG islands using a methylated DNA binding column
MEDLINE	Nat. Genet. 6 (3), 236-244 (1994)
PUBMED	94282070
COMMENT	8012384 Vector: pGEM-5zf(-) Resources are available from the UK MRC Human Genome Mapping Project Resource Centre, Hinxton, Cambridgeshire CB10 1RQ, UK. See URL: http://www.hgmp.mrc.ac.uk/ for details or contact: bihelp@hgmp.mrc.ac.uk.
FEATURES	Location/Qualifiers 1..218 /organism="Homo sapiens" /db_xref="taxon:9606" /clone="53g8" /sex="male" /tissue_type="blood" /clone_lib="CGI-1" /dev_stage="adult"
BASE COUNT	57 a 43 c 85 g 31 t 2 others
ORIGIN	
Query Match	4.9%; Score 186.8; DB 9; Length 218;
Best Local Similarity	96.8%; Pred. No. 3.3e-34;
Matches	211; Conservative 0; Mismatches 4; Indels 3; Gaps 2
QY	98 TAAGCGCGCTAGGACAAGG--AGAAAGGTGTATCTCGCGGCTGTGGCGCTGCACATAC 155 
Db	1 TAAGCGCGCTAGGACAAGGTAGTAAAGGTGTATCTCGCGGCTGTGGCGCTGCACATAC 60 
QY	156 AAGACCATCGCGCCAGCGTGGGATCAGCAGGAGGAAGCGCTGAGCGAGAGCGATGGGGAT 151 
Db	61 AAGACCATCGCGCCAGCGTGGGATCAGCAGGAGGAAGCGCTGAGCGAGAGCGATGGGGAT 151 
QY	216 GAGTAGCCATGGCTTCTCCGCGGAGAGCTCCCGGGGAGTAGAGTGGGAATATGAC 275 
Db	121 GAGTAGCCATGGCTTCTCCGCGGAGAGCTCCCGGGGAGTAGAGTGGGNANTATGAC 188 
QY	276 GAA-GAGGAGGAGAAACCAGCTGGAGATTGAGAGAC 312 
Db	181 GAAGGAGGAGGAGAAACCAGCTGGAGATTGAGAGAC 218 
RESULT 15	
G54201	
LOCUS	G54201 369 bp DNA linear STS 03-AUG-1996
DEFINITION	BI3G14/T7 Human Chromosome 12 Homo sapiens STS genomic clone
ACCESSION	RPC1-11-BI3G14 T7, sequence tagged site.
VERSION	G54201
KEYWORDS	G54201.1 GI:5690177
SOURCE	STS.
ORGANISM	Homo sapiens. Homo sapiens Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo. 1 (bases 1 to 369) Montgomery, K.T., Marondel, I., Kneitz, S., Yang, R., Lau, S., Lee, E., Miller, A., Issar, A., Shim, C., Decker, J., Chiu, D. and Kucherlapati, R. A Sequence Ready Map of Human Chromosome 12 (1999) Unpublished (1999)
REFERENCE	
AUTHORS	Contact: Raju Kucherlapati
TITLE	Kucherlapati Lab., Department of Molecular Genetics
JOURNAL	Albert Einstein College of Medicine (AECOM)
COMMENT	1300 Morris Park Avenue, Bronx, NY 10461, USA Tel: 7184302824 Fax: 7184308778 Email: Kucherla@ecom.yu.edu Primer A: AGCTGTCCACGTTCTCTC Primer B: GCTCTTTCCGTTGTCACG



GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 18, 2003, 13:47:23 ; Search time 46 Seconds  
(without alignments)  
4313.552 Million cell updates/sec

Title: US-09-763-985A-2  
Perfect score: 4994  
Sequence: 1 MATAAETSAPEAEKAGP.....AATEAPKMSNADFAKLEFLRK 963

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SPTREMBL\_21.\*

- 1: sp\_archaea.\*
- 2: sp\_bacteria.\*
- 3: sp\_fungi.\*
- 4: sp\_human.\*
- 5: sp\_invertebrate.\*
- 6: sp\_mammal.\*
- 7: sp\_mhc.\*
- 8: sp\_organelle.\*
- 9: sp\_phase.\*
- 10: sp\_plant.\*
- 11: sp\_rodent.\*
- 12: sp\_virus.\*
- 13: sp\_vertebrate.\*
- 14: sp\_unclassified.\*
- 15: sp\_rvirus.\*
- 16: sp\_bacteriap.\*
- 17: sp\_archaeap.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	4994	100.0	963	4	Q15020 homo sapien
2	4379.5	87.7	962	11	Q9JL18 mus musculus
3	1812	36.3	364	4	Q86J95
4	902	18.1	816	10	Q8S919
5	864	17.3	836	5	Q17430
6	823	16.5	768	10	Q9STW7
7	506.5	10.1	1014	3	Q9USY2
8	425	8.5	941	5	Q9WAD2
9	420	8.4	943	5	Q18352
10	349	7.0	826	3	Q01491
11	293.5	5.9	705	13	Q06459
12	280	5.6	744	5	Q16376
13	272	5.4	1305	5	Q9NFE2
14	263.5	5.3	665	10	Q9FJ37
15	252.5	5.1	715	11	Q9QZX1
16	249.5	5.0	707	11	Q99K50

17	248.5	5.0	610	10	Q9LIH8	Q9LIH8 arabidopsis
18	243	4.9	557	10	Q9FVQ1	Q9FVQ1 arabidopsis
19	241.5	4.8	1009	5	Q9VAT3	Q9VAT3 drosophila
20	238.5	4.8	960	4	Q8BPY6	Q8BPY6 homo sapien
21	236.5	4.7	673	10	Q9FNM3	Q9FNM3 arabidopsis
22	234	4.7	836	4	Q9NYD8	Q9NYD8 homo sapien
23	234	4.7	848	4	Q9BZJ1	Q9BZJ1 homo sapien
24	233.5	4.7	476	5	Q27199	Q27199 tetrahymena
25	232	4.6	690	11	Q9CQC1	Q9CQC1 mus musculus
26	231.5	4.6	724	3	Q9HF03	Q9HF03 cryptococcus
27	229.5	4.6	675	10	Q9LKS1	Q9LKS1 arabidopsis
28	227.5	4.6	571	3	Q8X003	Q8X003 neurospora
29	226	4.5	575	3	Q9Y7A8	Q9Y7A8 neurospora
30	226	4.5	687	4	Q8WY15	Q8WY15 homo sapien
31	225	4.5	687	4	Q9BZJ2	Q9BZJ2 homo sapien
32	224	4.5	836	4	Q9NOH5	Q9NOH5 homo sapien
33	222.5	4.5	599	10	Q8SD35	Q8SD35 arabidopsis
34	221.5	4.4	711	10	Q8S5Y7	Q8S5Y7 oryza sativ
35	221.5	4.4	1972	11	Q8R384	Q8R384 mus musculus
36	219.5	4.4	717	11	Q99L17	Q99L17 mus musculus
37	217	4.3	564	4	Q9GZW7	Q9GZW7 homo sapien
38	216.5	4.3	708	4	Q9GQD6	Q9GQD6 homo sapien
39	216.5	4.3	717	4	Q12996	Q12996 homo sapien
40	215	4.3	733	3	Q14233	Q14233 schizosacch
41	214.5	4.3	611	10	Q41042	Q41042 pisum sativ
42	212.5	4.3	524	4	Q14498	Q14498 homo sapien
43	212.5	4.3	594	5	Q9VM49	Q9VM49 drosophila
44	212.5	4.3	883	5	Q9V6S4	Q9V6S4 drosophila
45	211	4.2	495	10	Q9ASP6	Q9ASP6 arabidopsis

# ALIGNMENTS

## RESULT 1

Q15020	PRELIMINARY;	PRT;	963 AA.
ID	Q15020		
AC	Q15020;		
DC	Q1-NOV-1996 (TREMBLrel. 01, Created)		
DT	Q1-NOV-1996 (TREMBLrel. 01, Last sequence update)		
DT	Q1-JUN-2002 (TREMBLrel. 21, Last annotation update)		
DE	ORF.		
OS	Homo sapiens (Human).		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.		
OX	NCBI_TaxID=9606;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RX	MEDLINE=96127530; PubMed=8590280;		
RA	Nagase T., Seki N., Tanaka A., Ishikawa K., Nomura N.;		
RT	"Prediction of the coding sequences of unidentified human genes. IV. The coding sequences of 40 new genes (KIAA0121-KIAA0160) deduced by analysis of cDNA clones from human cell line KG-1.";		
RL	DNA Res. 2:167-174(1995).		
RN	[2]		
RP	SEQUENCE FROM N.A.		
RA	Itoh K., Yang D., Sasatomi T., Nakao M., Shichiho S., Takasu H.,		
RA	Matsumoto H., Mori K., Yamana H.;		
RT	"SART-3 (Squamous cell carcinoma antigen recognized by T cells).";		
RL	Submitted (DEC-1998) to the EMBL/Genbank/DBJ databases.		
DR	EMBL; D63879; BAA09929.1; -		
DR	EMBL; AB020880; BAA78384.1; -		
DR	HSSP; P09012; 2UIA.		
DR	InterPro; IPR003107; HAT.		
DR	InterPro; IPR000504; RNA_rec_mot.		
DR	Pfam; PF00076; rrm; 2.		
DR	SMART; SM00386; HAT; 6.		
DR	SMART; SM00360; RRM; 2.		
DR	PROSITE; PS01002; RRM; 2.		
DR	PROSITE; PS00030; RRM_RNP_1; UNKNOWN_2.		
SQ	SEQUENCE 963 AA; 109934 MW; 06526CEB8B819102A CRC64;		

Query Match 100.0%; Score 4994; DB 4; Length 963;

Best Local Similarity 100.0%; Pred. No. 2.le-291;		Matches 963; Conservative 0; Mismatches 0; Indels 0; Gaps 0;	
QY	1	MATAETSASEPEAEKAGPKADGDEDEKAAATRRKVLRSRAVAATAATYKTMGPWQOEE	60
Db	1	MATAETSASEPEAEKAGPKADGDEDEKAAATRRKVLRSRAVAATAATYKTMGPWQOEE	60
QY	61	GVSESDGYAMASSAESSPGYEWYDEEEKNQLEIERLEBQLSINYVDYNCHVDLIR	120
Db	61	GVSESDGYAMASSAESSPGYEWYDEEEKNQLEIERLEBQLSINYVDYNCHVDLIR	120
QY	121	LLRLEGELTKVRMAROKMSEIPLTEELWLEWHLDEISMAQDGLDREHYVDLFEKAVDY	180
Db	121	LLRLEGELTKVRMAROKMSEIPLTEELWLEWHLDEISMAQDGLDREHYVDLFEKAVDY	180
QY	181	ICPNIWLEYGQYSGVGIGKGLKRVSRFERRALSSVGLHMTKGLALWEAYREFESAIYE	240
Db	181	ICPNIWLEYGQYSGVGIGKGLKRVSRFERRALSSVGLHMTKGLALWEAYREFESAIYE	240
QY	241	AARLEKVSFLRRQLAIPLYDMEATFAEYEWSESDPIPESVIQNYNKALQLEKYPYEE	300
Db	241	AARLEKVSFLRRQLAIPLYDMEATFAEYEWSESDPIPESVIQNYNKALQLEKYPYEE	300
QY	301	ALLOAEAPRLAEYQAYIDFEMKIGDPARTQIIFERALVENCLVPDLWIRYSQYLDRLQKV	360
Db	301	ALLOAEAPRLAEYQAYIDFEMKIGDPARTQIIFERALVENCLVPDLWIRYSQYLDRLQKV	360
QY	361	KDLVLSVHNRAINCPTWALVSRYLLAMERHGVHDHOVISVTFEKNALNAGFTQATDYVEI	420
Db	361	KDLVLSVHNRAINCPTWALVSRYLLAMERHGVHDHOVISVTFEKNALNAGFTQATDYVEI	420
QY	421	WQAYLDYLRVRVDFKODSKELEELRAAFTRALEYKQVEERFNEGSDPSCVIMQNWAR	480
Db	421	WQAYLDYLRVRVDFKODSKELEELRAAFTRALEYKQVEERFNEGSDPSCVIMQNWAR	480
QY	481	IEARLCNNMOKARELWDSIMTRGNAKYANMWLEYNLERAHGDTQCHRKALHRAVQCTSD	540
Db	481	IEARLCNNMOKARELWDSIMTRGNAKYANMWLEYNLERAHGDTQCHRKALHRAVQCTSD	540
QY	541	YPEHVCEVLLTWERTEGSLEDWDIAVQKTETRLARVNEQRMKAEEAALVQOEEKBAQ	600
Db	541	YPEHVCEVLLTWERTEGSLEDWDIAVQKTETRLARVNEQRMKAEEAALVQOEEKBAQ	600
QY	601	RRARAEEKALKKKKIRGPEKRGADDEDEKWDGDEDEEQPSKRRRVENSIPAAGETQNV	660
Db	601	RRARAEEKALKKKKIRGPEKRGADDEDEKWDGDEDEEQPSKRRRVENSIPAAGETQNV	660
QY	661	EVAAGPAGKCAAVDVEPPSKOKEKAAKSLKRDMPKVLHDSKDSITVFNLPYSMQEPDT	720
Db	661	EVAAGPAGKCAAVDVEPPSKOKEKAAKSLKRDMPKVLHDSKDSITVFNLPYSMQEPDT	720
QY	721	KLRPLFEACGEVYQIRPIEFSNRGDFRGYCYVEFKESALQALEMDRKSVGRPMFVSPC	780
Db	721	KLRPLFEACGEVYQIRPIEFSNRGDFRGYCYVEFKESALQALEMDRKSVGRPMFVSPC	780
QY	781	VKSKNPDPKVFYRSTSLKHKLFISGLPFSCTKEELEICKAHGTVDKDLRLVTNRAGRP	840
Db	781	VKSKNPDPKVFYRSTSLKHKLFISGLPFSCTKEELEICKAHGTVDKDLRLVTNRAGRP	840
QY	841	KGLAYVEYENESQASQAVMKMGMTIKENIKVATSNPPQKRVPEKPETRKAPGPMLLP	900
Db	841	KGLAYVEYENESQASQAVMKMGMTIKENIKVATSNPPQKRVPEKPETRKAPGPMLLP	900
QY	901	QTYGARGKRTQLSLLPRALQPSAAAPQAEANGPAAAPAAVAPAAEPKMSNADFAKLF	960
Db	901	QTYGARGKRTQLSLLPRALQPSAAAPQAEANGPAAAPAAVAPAAEPKMSNADFAKLF	960
QY	961	LRR 963	
Db	961	LRR 963	

RESULT 2

Q9JLI8

ID	Q9JLI8	PRELIMINARY;	PRT;	962 AA.
AC	Q9JLI8;			
DT	01-OCT-2000 (TrEMBLrel. 15, Created)			
DT	01-OCT-2000 (TrEMBLrel. 15, Last sequence update)			
DT	01-JUN-2002 (TrEMBLrel. 21, Last annotation update)			
DE	Tumor-rejection antigen SART3.			
GN	SART3.			
OS	Mus musculus (Mouse).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.			
OX	NCBI_TaxID=10090;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=BALB/C; TISSUE=SQUAMOUS CELL CARCINOMA;			
RX	MEDLINE=20222349; PubMed=10761712;			
RA	Harada K., Yamada A., Mine T., Kawagoe N., Takasu H., Itoh K.;			
RT	"Mouse homolog of the human SART3 gene encoding tumor-rejection antigen.";			
RL	Jpn. J. Cancer Res. 91:239-247(2000).			
DR	EMBL; AF172722; AAF65228.1; -			
DR	HSSP; P09012; 2U1A.			
DR	MGI; MGI:1858230; Sart3.			
DR	InterPro; IPR003107; HAT.			
DR	InterPro; IPR000504; RNA_rec_mot.			
DR	Pfam; PF00076; rtm; 2.			
DR	SMART; SM00386; HAT; 6.			
DR	SMART; SM00360; RRM; 2.			
DR	PROSITE; PS0102; RRM; 2.			
DR	PROSITE; PS00030; RRM_RNP_1; UNKNOWN_2.			
SQ	SEQUENCE 962 AA; 109618 MW; 23HC235125E7A09C CRC64;			

Query Match 87.7%; Score 4379.5; DB 11; Length 962;  
Best Local Similarity 87.0%; Pred. No. 1.6e-254;  
Matches 839; Conservative. 52; Mismatches 70; Indels 3; Gaps 3;

QY	1	MATAETSASEPEAEKAGPKADGDEDEKAAATRRKVLRSRAVAATAATYKTMGPWQOEE	60
Db	1	MATAETSASEPEAEKAGPKADGDEDEKAAATRRKVLRSRAVAATAATYKTMGPWQOEE	60
QY	61	GVSESDGY-AMASSAESSPGYEWYDEEEKNQLEIERLEEQLSINYVDYNCHVDLI	119
Db	61	GVSESDGY-AMASSAESSPGYEWYDEEEKNQLEIERLEEQLSINYVDYNCHVDLI	120
QY	120	LLRLEGELTKVRMAROKMSEIPLTEELWLEWHLDEISMAQDGLDREHYVDLFEKAVKD	179
Db	121	LLRLEGELTKVRMAROKMSEIPLTEELWLEWHLDEISMAQDGLDREHYVDLFEKAVKD	180
QY	180	YICPNIWLEYGQYSGVGIGKGLKRVSRFERRALSSVGLHMTKGLALWEAYREFESAIY	239
Db	181	YICPNIWLEYGQYSGVGIGKGLKRVSRFERRALSSVGLHMTKGLALWEAYREFESAIY	240
QY	240	EAARLEKVSFLRRQLAIPLYDMEATFAEYEWSESDPIPESVIQNYNKALQLEKYPYEE	299
Db	241	EAARLEKVSFLRRQLAIPLYDMEATFAEYEWSESDPIPESVIQNYNKALQLEKYPYEE	300
QY	300	EALLOAEAPRLAEYQAYIDFEMKIGDPARTQIIFERALVENCLVPDLWIRYSQYLDRLQK	359
Db	301	EALLOAEAPRLAEYQAYIDFEMKIGDPARTQIIFERALVENCLVPDLWIRYSQYLDRLQK	360
QY	360	VKDLVLSVHNRAINCPTWALVSRYLLAMERHGVHDHOVISVTFEKNALNAGFTQATDYVE	419
Db	361	VKDLVLSVHNRAINCPTWALVSRYLLAMERHGVHDHOVISVTFEKNALNAGFTQATDYVE	420
QY	420	WQAYLDYLRVRVDFKODSKELEELRAAFTRALEYKQVEERFNEGSDPSCVIMQNW	479
Db	421	WQAYLDYLRVRVDFKODSKELEELRAAFTRALEYKQVEERFNEGSDPSCVIMQNW	480
QY	480	RTEARLCNNMOKARELWDSIMTRGNAKYANMWLEYNLERAHGDTQCHRKALHRAVQCT	539
Db	481	RTEARLCNNMOKARELWDSIMTRGNAKYANMWLEYNLERAHGDTQCHRKALHRAVQCT	540
QY	540	DYPEHVCEVLLTWERTEGSLEDWDIAVQKTETRLARVNEQRMKAEEAALVQOEEKAE	599

```
Db 541 DYPEHVCVLLTMEETGLTDWDLAIQKTRLARVNEQRMKAKEAALVQOEEKAE 600
Qy 600 QRKARAKKALKKKKKIRGPEKRCAGDEDEKWKGDDEEOPSKRRRVENSIPRAGETQN 659
Db 601 QRKVRABKALKKKKKTRGADKRREDEENGEWEEBOPSKRRRTENSIL-ASGEASA 659
Qy 660 VEVAAGPAGKCAAVDVEPPSKOKEKAAKSLKRDMPKVLHDSKSDITTFVSNLPSMQEPD 719
Db 660 MKEETELSGKCLTIDVGPSPKOKEKAAKSLKRDMPKVAHDSKSDITTFVSNLPSIEPE 719
Qy 720 TKLRPLFACCEGVQIRIFNRRGDFRGYCVVEKEEKSALQALEMDRKSVEGRPMFVSP 779
Db 720 VKLRPLFVCEGVQIRIFNRRGDFRGYCVVEKEEKSALQALEMDRKSVEGRPMFVSP 779
Qy 780 CVDSKSNPDFKVFYRSTSLKHLFISGLPFSCTKEELEECAGTGVKDLRLVNTNRAGK 839
Db 780 CVDSKSNPDFKVFYRSTSLKHLFISGLPFSCTKEELEECAGTGVKDLRLVNTNRAGK 839
Qy 840 PKGLAYVEYENESQASQAVKMDGMTIKENIKVAISNPPQKVPKPEKTRKAPGPMML 899
Db 840 PKGLAYVEYENESQASQAVKMDGMTIRENVIKVAISNPPQKVPKPEKTRKAPGPMML 899
Qy 900 POTYGARCKGTQISLLPRALQPSAAPQAPGPAAPAAAPAAAPAAAPAAAPAAAPAA 959
Db 900 RQMGYARGKGTQISLLPRALQOR-OGAAPQAPGPAAPGPAAPVAPVAPVAPVAPVAPV 958
Qy 960 FLRK 963
Db 959 LLRK 962

RESULT 3
Q96J95
ID Q96J95 PRELIMINARY; PRT; 364 AA.
AC Q96J95;
DT 01-DEC-2001 (Tremblrel. 19, Created)
DT 01-DEC-2001 (Tremblrel. 19, Last sequence update)
DE KIA0156 Isoform.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=BRAIN;
RA Liu Y., He J.J.;
RT "Cloning and characterization of a novel Tat-binding protein.";
RL Submitted (MAY-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF387506; AAK69347.1;
SQ SEQUENCE 364 AA; 41828 MW; D86B2562A7FA2E4E CRC64;

Query Match 36.3%; Score 1812; DB 4; Length 364;
Best Local Similarity 100.0%; Pred. No. 5.2e-101;
Matches 350; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MATAETSASPEAESKAGPADGEDEVKAARTRRKVLRSRAVAATYKTMGPWDQOEE 60
Db 1 MATAETSASPEAESKAGPADGEDEVKAARTRRKVLRSRAVAATYKTMGPWDQOEE 60

Qy 61 GVSESDGDEYAMASSAESSPGYEWYDEEENKQLETERLEEQSLINVDYNCHVDLIR 120
Db 61 GVSESDGDEYAMASSAESSPGYEWYDEEENKQLETERLEEQSLINVDYNCHVDLIR 120

Qy 121 LLRLEGELTKVRMAROKMSEIFPLTEELWLEWLHDEISMAQDGLDREHVDLFEKAVDY 180
Db 121 LLRLEGELTKVRMAROKMSEIFPLTEELWLEWLHDEISMAQDGLDREHVDLFEKAVDY 180

Qy 181 ICPNIWLEYGOYVSGIGQKGLKVRVSVFERRALSSVGLHMTKGLALWEAREFESAVE 240
Db 181 ICPNIWLEYGOYVSGIGQKGLKVRVSVFERRALSSVGLHMTKGLALWEAREFESAVE 240
```

```
Qy 241 AARLEKHSLFRRLQALPIYDMEATFAEYEWSEDPESVYQNTYKALQLEKYPYEE 300
Db 241 AARLEKHSLFRRLQALPIYDMEATFAEYEWSEDPESVYQNTYKALQLEKYPYEE 300

Qy 301 ALLOAEAPRLAYQAYIDFEMKIGDPARIQILIFERALVENCILVPLDIWIRY 350
Db 301 ALLOAEAPRLAYQAYIDFEMKIGDPARIQILIFERALVENCILVPLDIWIRY 350

RESULT 4
Q8S919
ID Q8S919 PRELIMINARY; PRT; 816 AA.
AC Q8S919;
DT 01-JUN-2002 (Tremblrel. 21, Created)
DT 01-JUN-2002 (Tremblrel. 21, Last sequence update)
DE ATG24270/T2A6_100.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsi.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RA Shinn P., Chen H., Cheuk R., Kim C.J., Meyers M.C., Banh J.,
RA Bowser L., Carninci P., Chang E., Dale J.M., Goldsmith A.D.,
RA Hayashizaki Y., Ishida J., Jones T., Kamiya A., Karlin-Neumann G.,
RA Kawai J., Lam B., Lee J.M., Lin J., Miranda M., Narusaka M.,
RA Nguyen M., Onodera C.S., Palm C.J., Quach H.L., Sakurai T., Satou M.,
RA Seki M., Southwick A., Tang C.C., Toriumi M., Wu H.C., Yamada K.,
RA Yamamura Y., Yu G., Yu S., Shinzaki K., Davis R.W., Theologis A.,
RA Ecker J.R.;
RT "Arabidopsis cDNA clones.";
RL Submitted (JAN-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AY075670; AAL77677.1;
SQ SEQUENCE 816 AA; 92808 MW; 2260EDC78469DBAF CRC64;

Query Match 18.1%; Score 902; DB 10; Length 816;
Best Local Similarity 28.3%; Pred. No. 6.5e-46;
Matches 271; Conservative 154; Mismatches 328; Indels 206; Gaps 34;

Qy 56 DOQEGVSESDGDEYAMASSAESSPGYEWYDEEENKQLETERLEEQSLINVDYNCH 115
Db 13 DQKMEASAPARADPPSDSDS--GSDSDSEDEAESNH-QIVTLESELSANPYNDAY 69

Qy 116 VDLRLLEGELTKVRMAROKMSEIFPLTEELWLEWLHDEISMAQDGLDREHVDY--L 172
Db 70 VOYIKLLRKTANLEKLRQAREAMSAIFPLSPSLWLEWARDEASLAAS---ENVPEIVML 125

Qy 173 FEKAVKDYICPNWLEYGOY-----SVGGIGQKGLKVRVSVFERRALSSVGLHMTKGLA 226
Db 126 YERGLSDYQSVSLWCDYLSFMLEFDPSPVGRYPSE-GISKMSLFEAIPAAGFHVTEGTR 184

Qy 227 LWAEYREFESATV-----EA-----ARLEKHSLFRRLQALPIYDMEATFAEYEW 272
Db 185 IWEGYREFEQGVLATIDEADIEERNKQIRINSIFRHLSPLESLSSLIAYKATWELEQ 244

Qy 273 -----SED--PIPESVYQNTYKALQLEKYPYEEALQ---AEAPRLAYQAYIDPE 320
Db 245 GIDLDIGSDLSKVSQSHQAVANKKAQMYSERAHLEENISKODLSDETEKFEQFMNIRKE 304

Qy 321 MKIGDPARIQILIFERALVENCILVPLDIWIRYSOYLDRLQKLVKDLVLSINRNPWTVA 380
Db 305 KTSQDPTRVQAIYERAVAEYPVSSDLWDITYVLDKTLKVGKAIHAYSRATRSCPTGCD 364

Qy 381 LWSRYLLAMERHGVHQVISTVFEKALNAGFTQATYVEIWOAYLDYLRVRVDFKODSSK 440
Db 365 LWARYLLALERSASEKEIYDVFEKSLQCTESSFEYLDLTLRVLDGLRRM----- 416

Qy 441 ELEEELRAATRALEYLKQVEERFNEGDPSCVIMON-----WARTEARLCNNMQ 490
Db 417 ----LSTRMLEALDY--SLIRETFQASDYLTPHMONTDLSLHLHTYWANLELNKGDLA 470
```

QY 491 KARELWDSIMTGNKAKYANMWLEYNNLERAHGDTQHCKALHRAVQCTSD--YPHERVCVEP 548  
 Db :  
 QY 471 GARGVWDSPFLKXSGGMLA-AWHAVIDMEVHLGHKEARSYYRCYTTRKKFDGTGSDEICKG 529  
 Db :  
 QY 549 LITMERTBGSLEDWDIAVOKTETRLARVNQRMKAAEKAALVQQBEERAEQAQRKARAERK 608  
 Db :  
 QY 530 WURFEREHGDLEHFDLAVQKVMPRL-----EELQLMRLOQEESTPVKPSAGLUKE 577  
 Db :  
 QY 609 KALKKKKIRPEKRGADDEDKEKWGEDDEEQPSKR--RRVENSIPAAGETONVEVAAGP 666  
 Db :  
 QY 578 HSSQRRK-----AEQ-----NVEESLAKRKRKSQEKVEDLGGQSATV----P 616  
 Db :  
 QY 667 AGK-CAADVDPVPKQKEPKASLRKMDPKVLHDSSKDSITVFVSNIPLYSMQBPDTKLRLP 725  
 Db :  
 QY 617 ATKNVKAENGTAQSDKEETEDVKPLPKPYV---RDCEAFATISNLVSAQBED--IRKF 670  
 Db :  
 QY 726 FEACGEVVQIIRPI-FSNRNGDFRGYCYVEEFEEKKSALAQLEMDRKSVEGRPFMFVPCVDKS 784  
 Db :  
 QY 671 FDDGGGVDSIRILLHHKDTGKPRGLAYADFVDDHELAALAAIKNRKMFFGKKISIA----- 725  
 Db :  
 QY 785 KNPDFKVFRYSTSLSEKHKLFTISGLFPSCFCTKEELEEICKAHGTVKDRLRLVTNRAGKPKGLA 844  
 Db :  
 QY 726 SNP-----KKGK-----KEFTRRNGDGSNKDPSLISEKAKAPLG-- 761  
 Db :  
 QY 845 YVEENESQASQAVNMKDMGTWIKENIIKVAISNPPOKRVPEKPETRKAPGGPMLLPDQTGY 904  
 Db :  
 QY 762 -----GETEGER-----KGNEVEV----- 775  
 Db :  
 QY 905 ARGKGRTQLSLLPALORPSAAAQAENGPAAPAAVAPAATAEPKMSNADFAKLFRLK 963  
 Db :  
 QY 776 R-GK---NTFEAVRNVRKPLGYTTPK-----PSADETPK-SNDEFNRMFLKK 816  
 Db :

RESULT 5  
 Q17430 PRELIMINARY; PRT; 836 AA.  
 ID Q17430 AC Q17430; DT 01-JAN-1999 (TREMBLrel. 09, Created)  
 DT 01-JAN-1999 (TREMBLrel. 09, Last sequence update)  
 DE 01-JUN-2002 (TREMBLrel. 21, Last annotation update)  
 DE Hypothetical 95.5 kDa protein B0035.12 in chromosome IV.  
 GN B0035.12.  
 OS Caenorhabditis elegans.  
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditiida; Rhabditoidea;  
 OC Rhabditiidae; Pelodierinae; Caenorhabditis.  
 NCBI\_Taxid=6239;  
 RN [1]  
 RC SEQUENCE FROM N.A.  
 RP STRAIN-BRISTOL N2;  
 RL White S.;  
 RA Submitted (MAY-1996) to the EMBL/GenBank/DDBJ databases.  
 RI -!- SIMILARITY: WEAK, TO HUMAN KIAA0156.  
 DR EMBL; 273102; CAA97405.1;  
 DR WormPep; B0035.12; CE05167.  
 DR InterPro; IPR003107; RNA.  
 DR InterPro; IPR000504; RNA\_rec\_mot.  
 DR Pfam; PF00076; rrm; 2.  
 DR SMART; SM00386; HAT; 3.  
 DR SMART; SM00360; RRM; 2.  
 DR PROSITE; PS50102; RRM; 2.  
 DR PROSITE; PS00030; RRM\_RNP\_1; UNKNOWN\_2.  
 KW Hypothetical protein.  
 SQ SEQUENCE 836 AA; 95486 MW; 99912BF7436C0D2 CRC64;

## RESULT 6

09STW7

Q951W/ ID Q9STW7 PRELIMINARY; PRT; 768 AA.

Q95TW1; FRELIMNARI; 700 AA.  
LD  
AC  
Q95TW1;  
01-MAY-2000 (TrEMBLrel. 13, Created)  
DT  
DT  
01-MAY-2000 (TrEMBLrel. 13, Last sequence update)  
DT  
DT  
01-JUN-2002 (TrEMBLrel. 21, Last annotation update)  
DE  
DE  
Hypothetical 87.2 kDa protein.  
GN  
T22A6.100 OR A74G24270.  
OS  
Arabisopsis thaliana (Mouse-ear cress).  
OC  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
OC  
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
OC  
eurosids II; Brassicales; Brassicaceae; Arabidopsids.

[illegible]

```
Db 193 EIVFKNKIYDFHIDRLKLPHEQTEFTSLSTFTVNNWSPSEY-EDVMVKSNKRYETTLK 251
QY 295 --YKPEEALLOAEAPLAEYOAYIDFEMKIGDPA-----IQLIFERALVENCPLV 344
Db 252 RRAKFNKELLNSANHSLE--AYMDL-----INDESRRAEALQYITTLTLEIAVLPLIP 306
QY 345 DLWIRYSOYDLQRLKVLVSHNRRAIRNCPTWVALVSRYLILAMERHGVHQVIVSTPE 404
Db 307 ELWLQYTWLMSKVDFSSQASSVAERATRNCSWIGRIWSIKLTYMTLSGASISVACEED 366
QY 405 KALNAGFIQATDYVETWQAYLDLRRVDFKODSSKELEELRAAFTRALEYLKQVEERF 464
Db 367 RCLNSNLL--VNFDEVIDFSGFLKACLYLSSNEKDPQEFLLKHQIHKVEDYLK----- 418
QY 465 NESGDPSCVMQNAWIEARL-----CANNOKARELWDSITMTRGNKAKYANMWLE--- 513
Db 419 NKGSKD-----AMRIELSKIYLYSISFESVEKWC--SDMFHDFQNALYISRYI 470
QY 514 ---YNNLRAHGDQCRKALHRAVQCTSDYPHVCEVLLTMTERTSGSLEDWDIAVQKTE 570
Db 471 STMKNYPELA---AETLKSLYKNV---DQDQ-----LLQFYQSINDLNDCFTNS 517
QY 571 TRLARVNEORM---RAAEKEAALVQOEKEAKORRARAEEKALKKKKIRGPEKGADE 627
Db 518 HLYDVLNAQRISFKRQDLSFAETKQTVENTPLKVPQADDTAALSKKRKPGQE----- 571
QY 628 DDEKEKGD--DEEOPSKRRR-----VENSIPAGETONVEVAAGPAGKCAADVPEPS 679
Db 572 -----GDVFKSKPIEQHNRRELTVLVNLPSDISENELKIFFKDCGNIRIFILEDN 625
QY 680 KQKKAASLK-----RMPKVL-HDSSK---SITVFVSNLPSYMQEPTDKLR 723
Db 626 QKDVKVAQIEFSETSEVLAARTLKSIRGHSIQIHVDVTNLYVTNPTTYDELD--IT 683
QY 724 PLFEACGEVVOIRPFSNRGDFRGYCYFEKEKSALQALEMDRKSVEGR---PMFVSPC 780
Db 684 KLFSAGNVVVDVRRFSLRYNTNRRFCYVQMRKPDDEAHNALQHLKLEEKYPIQVFIS-- 741
QY 781 VDKSNKPDKPFYRSTSLKHLKIFSGLPFCSTKCELEICKAHGTVDKDLRLVTVNRAGK 840
Db 742 -----DPLRTPRSGAVYEGRELYNTNIDFKYNEKDVETFFRDYGOVESVR-IPKRFNH 795
QY 841 KGLAVYENESQASQAVMKMDGMTIKENIKVAISNP 878
Db 796 KGFYGVVMTNQDAENA--LSAAGKQLGNRVNLNWLSPK 832
[1]
RESULT 8
Q9W4D2 PRELIMINARY; PRT; 941 AA.
AC Q9W4D2;
DT 01-MAY-2000 (TREMELrel. 13, Created)
DT 01-MAY-2000 (TREMELrel. 13, Last sequence update)
DT 01-JUN-2002 (TREMELrel. 21, Last annotation update)
DE RNPAF protein (LD32008P).
GN RNPAF OR CG3312.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_Taxid=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BERKELEY;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Ananatiades P.G., Scher S.E., Li P.W., Hoskins R.A., Galie R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abril J.F., Agbayani A., An H.-J., Andrews-Frankkoch C., Baldwin D.,
```

```
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brotter P.,
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlike C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Foster C., Gabriellian A.E., Gaig N.S., Gelbart W.M., Glasser K.,
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Helman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert X., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kianos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wassarman D.A., Weinstein G.M., Weissenbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster."
RL Science 287:2185-2193(2000).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=BERKELEY;
RA Stapleton M., Brokstein P., Hong L., Agbayani A., Carlson J.,
RA Champe M., Chavez C., Dorsett V., Farfan D., Frise E., George R.,
RA Gonzalez M., Guarin H., Li P., Liao G., Miranda A., Mungall C.J.,
RA Nunoo J., Pacleb J., Paragas V., Park S., Phouanavong S., Wan K.,
RA Yu C., Lewis S.E., Rubin G.M., Celniker S.;
RL Submitted (Oct-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF003433; AAF46022.1;
DR EMBL: AY061395; AAL28943.1;
DR HSSP: P09012; IURN.
DR FlyBase: FBgn0014024; Rnp4F.
DR InterPro: IPR003107; RNA.
DR InterPro: IPR000504; RNA_rec_mot.
DR Pfam: PF00076; rtm; 1.
DR SMART: SM00386; HAT; 3.
DR PROSITE: PS00360; RRM; 1.
DR PROSITE: PS00030; RRM; 1.
DR PROSITE: PS00030; RRM; 1.
SQ SEQUENCE 941 AA; 106739 MW; EAF2775B499FECFD CRC64;
Query Match 8.5%; Score 425; DB 5; Length 941;
Best Local Similarity 20.3%; Pred. No. 3.4e-17;
Matches 221; Conservative 151; Mismatches 347; Indels 372; Gaps 40;
QY 7 TSASPEAESKAGPKADGEDEVKAAARTRRKYLRSRAVAAATYKTGPMADQOEVSSED 66
Db 87 SSDDEPSVEETGGNAGR-----GRANDSSSSDDVGVIE--GSELESNEVSDDSD 138
QY 67 GDEYAMASSAESPGYEWEYDEEEKNOLETERLEEQLSINVDYDNCVHDLRLRLLEG 126
Db 139 SD-----SDNAGGNQLERSY---OELNALPSKKFAQWVS-----LIGIAFKLN 179
QY 127 ELTKVMARQKMSKSEIFPLTEELWLEWHLDEISMAQDGLDREHYDLFEKAVKDYICPNW 186
Db 180 DLEKTESSVLEQLONLATVPAHVLKARLVVYTTQDEERKAFPEOCALGYYIS---- 235
QY 187 LEYGOYSGVIGKQKGLKVRSVFRALSSVGLHMTKGLALWEAVREFESAIVARLEK 246
Db 236 IPLSEYVNVYLVDQGNVQN-----HV-----LW-----AKLLA 263
```





```
Db 699 -----SPSKVK----- 705
QY 723 RPLFEACGEVQIRPIFSNRGDFRGYCYVEFEKESALQALEMDRKSVEGRPMFVSPCVD 782
Db 706 -----GPAEAESKN----- 716
QY 783 KSNPDKFVRYSTSLKHKLFISGLPFSCTKEELEICKAHGTVKDLRLVNRAGKPKG 842
Db 717 -----FKYSPNMEINKIFVRLNHPACSKRELHELFPFGTICKDRLVHLKLNQFKG 767
QY 843 LAVVEENESOAQVMKMGMTIKENIKVAISNPPQKVPE-KPETKAPCGPMLLPQ 901
Db 768 IAYVEFEKGEAQRVAGRGDGLFKMNISVALSNPPRGTSVAKPSV--AP----- 817
QY 902 TYGARGKRTQLSLLPRALQPSAAAPQ-----AENGPAAPAAVAAPAAAT----- 946
Db 818 -----KRRVPTSLIPTLVROEVAALKRLLLPEPGDISSTASVDVAIKREANGEEOK 871
QY 947 -----EAPK-----M 951
Db 872 GDVQERDEQGEQKGEQGEQKGEQKGEQKGEQKGEQKGEQKGEQKGEQKGEQKGEQK 931
QY 952 SNADFAKFLR 962
Db 932 SNDDFKLFLK 942

RESULT 10
Q01491
ID Q01491 PRELIMINARY; PRT; 826 AA.
AC Q01491
DT 01-NOV-1996 (TReMBLrel. 01, Created)
DT 01-NOV-1996 (TReMBLrel. 01, Last sequence update)
DT 01-JUN-2002 (TReMBLrel. 21, Last annotation update)
DE Colony 1.
GN COL1.
OS Ophiostoma ulmi (Dutch elm disease fungus).
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
OC Ophiostomatales; Ophiostomataceae; Ophiostoma.
OX NCBI_TaxID=5174;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MH75;
RA Royer J.C.; Hintz W.E.; Bowden C.; Kokurewicz K.; Li A.; Hubbes M.;
RL Submitted (SEP-1995) to the EMBL/GenBank/DBJ databases.
DR EMBL; U35661; AAA76605.1; -.
DR InterPro; IPR000504; RNA_rec_mot.
DR Pfam; PF00076; rrm; 4.
DR SMART; SM00360; RRM; 4.
DR PROSITE; PS00102; RRM; 4.
DR PROSITE; PS00030; RRM_RNP_1; UNKNOWN_1.
SQ SEQUENCE 826 AA; 90415 MW; 9C19CCB96B67EP46 CRC64;

Query Match 7.0%; Score 349; DB 3; Length 826;
Best Local Similarity 22.0%; Pred. No. 1e-12; Indels 264; Gaps 31;
Matches 178; Conservative 100; Mismatches 266;

QY 389 MERHGVDHVISVTEFKALNAGFIQATDYVEIQAQYLDYLRVRVDFKDSKSELELR-A 447
Db 1 MER--IKHAATNSLDR-----DGMVAVDMYATWCGYL-----RCAMDPNASDESVDLADS 51
QY 448 ATTRALEYLKQVEERFENES--GDFSCVIMQNWRIEARLCNN-----QKARELWDS 498
Db 52 GLVAALAEQVWGERRFGSAFOGDPN-----YRLERIMIVHLTEKHGAIDEAREHWDK 104
QY 499 INTRG-NKAYANWLEYI-----NLEAHGDT-----QHCRAALHRAVOC 537
Db 105 LAQKELLANDYFWLISYWMENMLLOSQKGTGRSPTPAPAKLSRTPSPASILQRAQV 164
QY 538 TS-DYPEHVEYVLLTMEREGSLEDWDIAVQKTTET--RLARVNEQRMKAEEKAALVOQE 594
Db 165 SOLNPWPVIEIYV---KHCNDFESSDVLQNALDEVHNLORVIAQRKDA--TAAQTAGA 219
```

```
QY 595 EKAQQRKARAEKAKKKKKIRGPEKRGADDD-----EKDW-----GD----- 635
Db 220 EAHARAYESQALEASAAAAEAQALGQQTGDPDQDLNAGVKRWKEAIASGETESATKPK 279
QY 636 -----DEEQPSKRRRVENSIPAA-----GETQNV 660
Db 280 KNEDICGVNAAAPNSQSQENQPAKRDRENTSVFVSNLPSDVTITKVRQYFREYGHVNI 339
QY 661 EYAAGPAGCAAVDVPPPSKQKKAASLRDMPKVLHD-----SSKDS--ITVFVSNLPYSM 715
Db 340 QLAHENGKSTVALVEFRSVEDAQTA-LIRD-GKYFGDHTISVKEAGITLYVNPFPST 397
QY 716 QEPDTKLRLPLFEACGEVQIRPIFSNRGDFRGYCYVEFEKESALQALEMDRKSVEGRPM 775
Db 398 D--DESLHRLFGKSSNIFGIRWPSLKYNAHRRFCYVFRDAESAATGLHGMLEGKYK 455
QY 776 FVSPCVDKSKNPDFKVFYRSTSLKHKLFISGLPFSCTKEELEICKAHGTVKDLRLVN 835
Db 456 LSVQYSD-----PAAKKAREGATDEGREVHIKNIPQEFDEKEIBQLVSKYGTVKRVRLHN 511
QY 836 RAGKPKGLAYVEENESQAQVMKMGMTIKENIKVAIS-----NPPQKVPEKPETR 890
Db 512 MAGRSRGSFAVDLETDEAERVVAELDKVLCTQILKVELSVSAKPKPSARETSVASESA 571
QY 891 KAPGGPMLLPQTYG-----ARGK----- 908
Db 572 TCGGGEAPDEAKGEAGNEARGNLHLSRFPALLGIPDVTNITRVRSLSAEPHGHTIKLKP 631
QY 909 -----GRTQSL-----LPRALQ----- 921
Db 632 EHGGAIIYEDETTAGKAOLALDSTOLEGHTLRVDAVPOLFEKESGVRIDRVDIARRPP 691
QY 922 RPSAAAPQAEANGPAAPAAVA----- 944
Db 692 KPTADTTSATMGKKAAPTAPALLPAFVRPVLGGKVGKHLGFSAISATTSKKGQPT 751
QY 945 ATE-----APKMSNADFAKFL 961
Db 752 ATNGAATSSSTSSDTAPKKSNAADFALFL 779

RESULT 11
Q06459
ID Q06459 PRELIMINARY; PRT; 705 AA.
AC Q06459;
DT 01-NOV-1996 (TReMBLrel. 01, Created)
DT 01-NOV-1996 (TReMBLrel. 01, Last sequence update)
DT 01-JUN-2002 (TReMBLrel. 21, Last annotation update)
DE Nucleolin.
OS Xenopus laevis (African clawed frog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipidea; Pipidae;
OC Xenopodinae; Xenopus.
OX NCBI_TaxID=8355;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=94039235; Pubmed=8223724;
RA Messmer B.; Dreyer C.;
RT "Requirements for nuclear translocation and nucleolar accumulation of
RT nucleolin of X. laevis.";
RL Eur. J. Cell Biol. 61:369-382(1993).
DR EMBL; X72957; CAA51460.1; -.
DR HSSP; P09651; LUPL.
DR InterPro; IPR000504; RNA_rec_mot.
DR Pfam; PF00076; rrm; 4.
DR SMART; SM00360; RRM; 4.
DR PROSITE; PS00102; RRM; 4.
DR PROSITE; PS00030; RRM_RNP_1; UNKNOWN_3.
SQ SEQUENCE 705 AA; 75521 MW; 8B2EC97F62403F02 CRC64;

Query Match 5.9%; Score 293.5; DB 13; Length 705;
Best Local Similarity 23.1%; Pred. No. 1.8e-09;
Matches 103; Conservative 71; Mismatches 147; Indels 125; Gaps 13;
```







**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 22, 2003, 06:20:31 ; Search time 530 seconds  
(without alignments)

16137.905 Million cell updates/sec

Title: US-09-763-985A-1

Perfect score: 3798

Sequence: 1 ccacgcgtccgctggcgact.....tcaaaaaaaaaaaaaaaaaa 3798

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 2185239 seqs, 1125999159 residues

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : N\_Geneseq\_101002.\*

1: /SID22/gcgdata/geneseq/geneseq-emb1/NA1980.DAT.\*  
2: /SID22/gcgdata/geneseq/geneseq-emb1/NA1981.DAT.\*  
3: /SID22/gcgdata/geneseq/geneseq-emb1/NA1982.DAT.\*  
4: /SID22/gcgdata/geneseq/geneseq-emb1/NA1983.DAT.\*  
5: /SID22/gcgdata/geneseq/geneseq-emb1/NA1984.DAT.\*  
6: /SID22/gcgdata/geneseq/geneseq-emb1/NA1985.DAT.\*  
7: /SID22/gcgdata/geneseq/geneseq-emb1/NA1986.DAT.\*  
8: /SID22/gcgdata/geneseq/geneseq-emb1/NA1987.DAT.\*  
9: /SID22/gcgdata/geneseq/geneseq-emb1/NA1988.DAT.\*  
10: /SID22/gcgdata/geneseq/geneseq-emb1/NA1989.DAT.\*  
11: /SID22/gcgdata/geneseq/geneseq-emb1/NA1990.DAT.\*  
12: /SID22/gcgdata/geneseq/geneseq-emb1/NA1991.DAT.\*  
13: /SID22/gcgdata/geneseq/geneseq-emb1/NA1992.DAT.\*  
14: /SID22/gcgdata/geneseq/geneseq-emb1/NA1993.DAT.\*  
15: /SID22/gcgdata/geneseq/geneseq-emb1/NA1994.DAT.\*  
16: /SID22/gcgdata/geneseq/geneseq-emb1/NA1995.DAT.\*  
17: /SID22/gcgdata/geneseq/geneseq-emb1/NA1996.DAT.\*  
18: /SID22/gcgdata/geneseq/geneseq-emb1/NA1997.DAT.\*  
19: /SID22/gcgdata/geneseq/geneseq-emb1/NA1998.DAT.\*  
20: /SID22/gcgdata/geneseq/geneseq-emb1/NA1999.DAT.\*  
21: /SID22/gcgdata/geneseq/geneseq-emb1/NA2000.DAT.\*  
22: /SID22/gcgdata/geneseq/geneseq-emb1/NA2001A.DAT.\*  
23: /SID22/gcgdata/geneseq/geneseq-emb1/NA2001B.DAT.\*  
24: /SID22/gcgdata/geneseq/geneseq-emb1/NA2002.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	3798	100.0	3798	21	AAZ90783
2	2886	76.0	3537	22	AAD21574
3	2563.2	67.5	3355	21	AAF16121
4	1331.2	35.1	1403	22	AAI98071
5	1056.4	27.8	32145	22	AAI04631
6	1056.4	27.8	32145	23	ABL97544
7	436	11.5	1344	23	AAS91136
8	398.2	10.5	437	21	AAA43952
9	393.4	10.4	432	24	ABL78750

10	392.4	10.3	430	24	ABL79050	Human ovarian canc
11	346.8	9.1	554	22	AAH98082	Murine 7-transmemb
c 12	314.4	8.3	543	22	AAK75790	Human immune/haema
c 13	310.4	8.2	312	22	ABA49131	Human breast cell
c 14	310.4	8.2	312	22	ABA67045	Human foetal liver
c 15	310.4	8.2	312	22	ABA34138	Probe #12604 for g
c 16	310.4	8.2	312	22	AAK15486	Human brain expres
c 17	310.4	8.2	312	22	AAK1215	Human bone marrow
c 18	310.4	8.2	312	22	AAI21971	Probe #11904 for g
c 19	310.4	8.2	312	22	AAI47256	Probe #15942 used t
c 20	310.4	8.2	312	22	AAI07663	Probe #7654 used t
c 21	310.4	8.2	312	24	ABS15226	Human genome-deriv
c 22	291.2	7.7	320	21	AAC01859	Human secreted pro
c 23	233.4	6.1	324	24	ABN77683	Human ORF2630 cDNA
c 24	219.8	5.8	475	22	ABA44036	Human breast cell
c 25	219.8	5.8	475	22	ABA54492	Human foetal liver
c 26	219.8	5.8	475	22	ABA42476	Probe #2742 for ge
c 27	219.8	5.8	475	22	AAK02781	Human brain expres
c 28	219.8	5.8	475	22	AAK28223	Human bone marrow
c 29	219.8	5.8	475	22	AAI12792	Probe #2725 for ge
c 30	219.8	5.8	475	22	AAI34142	Probe #2828 used t
c 31	219.8	5.8	475	22	AAI02705	Probe #2696 used t
c 32	219.8	5.8	475	24	ABS02736	Human genome-deriv
c 33	170	4.5	170	14	AAQ61082	Human brain Expres
c 34	136.4	3.6	381	20	AAV90234	EST clone DF950.
c 35	129.6	3.4	185	22	ABA47794	Human breast cell
c 36	129.6	3.4	185	22	ABA65686	Human foetal liver
c 37	129.6	3.4	185	22	ABA32772	Probe #11238 for g
c 38	129.6	3.4	185	22	AAK14089	Human brain expres
c 39	129.6	3.4	185	22	AAK39833	Human bone marrow
c 40	129.6	3.4	185	22	AAI20643	Probe #10576 for g
c 41	129.6	3.4	185	22	AAI45853	Probe #14539 used
c 42	129.6	3.4	185	22	AAI06342	Probe #6333 used t
c 43	129.6	3.4	185	24	ABS13928	Human genome-deriv
c 44	126.4	3.3	467	22	ABA42665	Human breast cell
c 45	126.4	3.3	467	22	ABA53092	Human foetal liver

#### ALIGNMENTS

#### RESULT 1

AAZ90783  
ID AAZ90783 standard; DNA; 3798 BP.  
XX  
AC AAZ90783;  
XX  
DT 23-JUN-2000 (first entry)  
XX  
DE Tumour antigen protein SART-3 encoding DNA.  
XX  
KW Tumour antigen protein; SART-3; cytotoxic T-cell; HLA antigen;  
XX  
KW tumour; ds.  
XX  
OS Homo sapiens.  
XX  
PN WO200012701-A1.  
XX  
PD 09-MAR-2000.  
XX  
PF 27-AUG-1999; 99WO-JP04622.  
XX  
PR 28-AUG-1998; 98JP-0242660.  
XX  
PA (SUMU ) SUMITOMO PHARM CO LTD.  
PA (ITOHI) ITOH K.  
XX  
PI Itoh K, Nakao M;  
XX  
DR WPI; 2000-237868/20.  
DR P-PSDB; AAY85422.  
XX  
PT Tumour antigen protein SART-3 recognized by cytotoxic T-cells binding to



PT HLA antigen for treatment and diagnosis of tumors

PS Claim 2; Page 58-63; 89pp.; Japanese.

XX This DNA encodes a tumour antigen protein (SART-3) which is recognized by  
CC cytotoxic T-cells binding to HLA antigen. SART-3 can be expressed by  
CC standard recombinant methodology. SART-3 and its peptide derivatives can  
CC be used in the prevention, treatment and diagnosis of tumours in vitro or  
XX in vivo.

SQ Sequence 3798 BP; 1036 A; 847 C; 1074 G; 841 T; 0 other;

Query Match 100.0%; Score 3798; DB 21.; Length 3798;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 3798; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY	1	CCACGCGTCCGATGGCGACTGCGCGCGAAGAACCTTCGCTTCAGAACCCGAGGCTGAGTCCA	60
DB	1		
QY	61	AGGCTGGGCCCAAGGCTGACGGAGAGGAGGATGAGGTTAAGGCGGCTAGGACAAGAGAA	120
DB	61		
QY	121	AGGTGTTATCGCGGGCTGTGGCGCTGCGACATACAAAGACCATGCGGGCCAGCGTGGGATC	180
DB	121		
QY	181	AGCAGGAGGAAGCGTGACGGAGAGCGATGGGGATGAGTACGCCATGGCTTCTCCGCGG	240
DB	181		
QY	241	AGAGCTCCCGGGGAGTACGAGTGGGAATATGACGAAGAGGAGGAGAGAAAAACCACTGG	300
DB	241		
QY	301	AGATTGAGAGACTGGAGGAGCAGTTGCTTATCAACGCTATGACTACAACTGCCATGTGG	360
DB	301		
QY	361	ACTTGATCAGACTGCTCAGGCTGGAGGGAGGCTTACCAAGTGAGGATGGCCGCCA	420
DB	361		
QY	421	AGATGAGTGAATCTTTCCCTTGACTGAAGAGCTCTGGCTGGAGTGGCTGCAATGACGAGA	480
DB	421		
QY	481	TCAGCATGCCAGGATGGCTGGACAGAGACGCTGATGACCTCTTTTGAGAAAGCCG	540
DB	481		
QY	541	TGAAGATTACATTTGCTCTACATTTGGCTAGAGTATGCCAGTACTCAGTTGGTGGGA	600
DB	541		
QY	601	TTGGTCAGAAAGTGCGCTTGAGAAAGTTCCGCTCGCTGTTGAAAGGGCTCTCTCGTCG	660
DB	601		
QY	661	TTGGTTTACATATGACCAAGGACTCGCCCTCTGGAGGCTTACCGAGAGTTTGAAGTG	720
DB	661		
QY	721	CGATTCTGAAGCTGCTCGGCTTGAGAAAGTCCACAGTCTTTTCGGGCGACAGTTGGCGA	780
DB	721		
QY	781	TCCCACTCTATGATATGGAGGCCACATTTGCAGAGTATGAAGAATGGTTCAGAAAGCCCAA	840
DB	781		
QY	841	TACCAGAGTCAGTAATTCAGAACTATACAAAGCAGCTACAGCAGCTGGAGAAATATAAAC	900
DB	841		

DB	841	TACCAGAGTCAGTAATTCAGAACTATACAAAGCAGCTACAGCAGCTGGAGAAATATAAAC	900
QY	901	CCATATGAAGACACACTGTTTGCAGGACAGGACCAAGGCTGCGAGATATCAAGCATATA	960
DB	901		
QY	961	TCGATTTTTCAGATGAAATTTGGCGATCCCTGCTCGCATTCAGTTGATCTTTGAGCGGCC	1020
DB	961		
QY	1021	TGGTCGAGAACTGCGCTTGTCCAGACTTATGATCCGTTACAGTACGTACCTAGATCGAC	1080
DB	1021		
QY	1081	AACTGAAACTAAAGGATTTGGTTTATCTGTACATAACCGCGCTATTAGAACTGCCCT	1140
DB	1081		
QY	1141	GGACAGTTCGCTTATGGAGTCGCTACCTTTGGCCATGGAGACATGGAGTTGATCATC	1200
DB	1141		
QY	1201	AAGTAATTTCTGTAACTTCGAGAAAGCTTTGAATGCCGGCTTCATCCAGGCCACTGATT	1260
DB	1201		
QY	1261	ATGTGAGATTTGGCAGGCATACCTTGATTACTTGAGGAGAGGGTTGATTCAAAACAG	1320
DB	1261		
QY	1321	ACTCCAGTAAAGAGCTGGAGAGTGGAGGCGGCTTTACTGTCGCTTTGGAGTATCTGA	1380
DB	1321		
QY	1381	ACAGAGGTGGAAGAGCGTTTCAATGAGAGTGGTGATCCAGCTCGGTGATTATGACAGA	1440
DB	1381		
QY	1441	ACTGGCTAGGATTTGAGGCTCGACTGTGCAATAACATGCAAGAGCTCGGAACTCTGG	1500
DB	1441		
QY	1501	ATAGCATCATGACAGAGGAAATGCCAAGTACGCCAACATGTGGCTAGAGTATTACAAC	1560
DB	1501		
QY	1561	TGGAAGAGCTCATGTTGACACCCAGCACTGCGGAGGCTCTGCACCGGCGCTCCAGT	1620
DB	1561		
QY	1621	GCACAGTACCTACCCAGAGCAGTCTGGAAGTGTACTCACCATTGAGAGGACAGAAAG	1680
DB	1621		
QY	1681	GTTCTTTAGAAGATTGGGATATAGCTGTTTTCAGAAAACTGAAACCCGATTAGCTCGTCA	1740
DB	1681		
QY	1741	ATGACAGAGAAATGAAGCTGCAGAGAGGAGAGAGCCCTTGTGCACGAGAGAGAGAA	1800
DB	1741		
QY	1801	AGGCTGAACCAACCGAAAAAGAGCTCGGGCTGAGAAAGAGCGTTTAAAAAGAGAGAA	1860
DB	1801		
QY	1861	TCAGAGGCCAGAGAGCGCGAGCAGATGAGGACGATGAGAAAGAGTGGGCGCATGATG	1920
DB	1861		
QY	1921	AGAGAGGAGCGCTTCCAAACCGCAGAGGGTGCAGAACAGCATCCCTGCAGCTGGAGAA	1980
DB	1921		

Db	1921	AGAAGAGCAGCCCTTCCAAACGACAGAGGGTCGAGAAACAGCATCCCTCGAGCTGGAGAA	1980
Qy	1981	CACAAAATGTAGAAGTAGCAGCAGGGCCCGCTGGGAAATGTGCTGCCGTAGATGTGGAGC	2040
Db	1981	CACAAAATGTAGAAGTAGCAGCAGGGCCCGCTGGGAAATGTGCTGCCGTAGATGTGGAGC	2040
Qy	2041	CCCTTTCGAAGCAGAGAGGAGGAGGAGCCCTCCCTGGAAGAGGGACATGCCCAAGGTGTGC	2100
Db	2041	CCCTTTCGAAGCAGAGAGGAGGAGGAGGAGCCCTCCCTGGAAGAGGGACATGCCCAAGGTGTGC	2100
Qy	2101	ACGACAGCAGCAAGGACAGCATCACCGTCTTTGTGCACACCACTGCCCTACAGCATGCAGG	2160
Db	2101	ACGACAGCAGCAAGGACAGCATCACCGTCTTTGTGCACACCACTGCCCTACAGCATGCAGG	2160
Qy	2161	AGCGGACAGCAAGCTCAGGCCACTCTTCGAGGCCCTGTGGGAGGTGTCAGATCCGAC	2220
Db	2161	AGCGGACAGCAAGCTCAGGCCACTCTTCGAGGCCCTGTGGGAGGTGTCAGATCCGAC	2220
Qy	2221	CCATCTTCAGCAACCGTGGGATTTCCGAGGTTACTGCTACGTGGAGTTTAAAGAGAGA	2280
Db	2221	CCATCTTCAGCAACCGTGGGATTTCCGAGGTTACTGCTACGTGGAGTTTAAAGAGAGA	2280
Qy	2281	AATCAGCCCTTCAGGCACTGGAGATGGACCGGAAAGTGTAGAGGGAGGCCCAATGTTG	2340
Db	2281	AATCAGCCCTTCAGGCACTGGAGATGGACCGGAAAGTGTAGAGGGAGGCCCAATGTTG	2340
Qy	2341	TTTCCCTCTGTGTGATGAAGCAAAACCCGATTTTAAGGTGTTTCAGGTACAGCACTT	2400
Db	2341	TTTCCCTCTGTGTGATGAAGCAAAACCCGATTTTAAGGTGTTTCAGGTACAGCACTT	2400
Qy	2401	CCCTAGAGAAACAAAGCTGTTCAATCAGCCCTGCTTCTCCTGTACTAAAGAGGAAC	2460
Db	2401	CCCTAGAGAAACAAAGCTGTTCAATCAGCCCTGCTTCTCCTGTACTAAAGAGGAAC	2460
Qy	2461	TAGAAGAAATCTGTAAGGCTCATGGCACCGTGAAGGACCTCAGGCTGTCACCAACCGG	2520
Db	2461	TAGAAGAAATCTGTAAGGCTCATGGCACCGTGAAGGACCTCAGGCTGTCACCAACCGG	2520
Qy	2521	CTGCAAAACCAAGGGCTGGCCCTACGTCAGTGGAGTATGAAATGAATCCAGCGCTCGCAGG	2580
Db	2521	CTGCAAAACCAAGGGCTGGCCCTACGTCAGTGGAGTATGAAATGAATCCAGCGCTCGCAGG	2580
Qy	2581	CTGTGATGAAGTAGGACGGATGACTATCAAGAGAACATCATCAAAAGTGGAATCAGCA	2640
Db	2581	CTGTGATGAAGTAGGACGGATGACTATCAAGAGAACATCATCAAAAGTGGAATCAGCA	2640
Qy	2641	ACCTCTCTCAGAGAAAGTTCAGAGAGCCAGAGACCAGGACCAAGGACCAAGTGGCCCA	2700
Db	2641	ACCTCTCTCAGAGAAAGTTCAGAGAGCCAGAGACCAGGACCAAGGACCAAGTGGCCCA	2700
Qy	2701	TGCTTTTCCGCGACATACGGAGCGAGGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG	2760
Db	2701	TGCTTTTCCGCGACATACGGAGCGAGGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG	2760
Qy	2761	CTGCTGCCCTGCAGCGCCCAAGTGTCTCAGCTCCTCAGGCTGAGAACGGCCCTGCCCGG	2820
Db	2761	CTGCTGCCCTGCAGCGCCCAAGTGTCTCAGCTCCTCAGGCTGAGAACGGCCCTGCCCGG	2820
Qy	2821	CTCTCTCAGTGTGCCGCCCGCCAGCAGCCACCGAGGACCCCAAGATGTCCTCAATGCCATTTG	2880
Db	2821	CTCTCTCAGTGTGCCGCCCGCCAGCAGCCACCGAGGACCCCAAGATGTCCTCAATGCCATTTG	2880
Qy	2881	CCAAGCTGTTTCTGAGAAAGTGAACGGGAGCGCTGGGAGACAGGAAATGCCTTACTTCACT	2940
Db	2881	CCAAGCTGTTTCTGAGAAAGTGAACGGGAGCGCTGGGAGACAGGAAATGCCTTACTTCACT	2940
Qy	2941	CTGCGCCGGGAGCTTCCACCAAGCAGCAGTGTGGGATGACAGGCGCTGGTGTGC	3000
Db	2941	CTGCGCCGGGAGCTTCCACCAAGCAGCAGTGTGGGATGACAGGCGCTGGTGTGC	3000
Qy	3001	TGCGTGTCTGCAACCAACAGATGGCTCTCGGCTTTACAGAGAGGGAGGGGTTCTAA	3060
Db	3001	TGCGTGTCTGCAACCAACAGATGGCTCTCGGCTTTACAGAGAGGGAGGGGTTCTAA	3060
Qy	3061	GTCAAGAGCCCTTTCAGTGTCTCCCTCATATTGAGGGCAGTGGCAGAAAAGTGACCACTCTG	3120
Db	3061	GTCAAGAGCCCTTTCAGTGTCTCCCTCATATTGAGGGCAGTGGCAGAAAAGTGACCACTCTG	3120
Qy	3121	CAGGCTGGGCCCCAGGATGTGGTGTCTGAGATAGTTTGTATCTTAAAGACTGAGGCACA	3180
Db	3121	CAGGCTGGGCCCCAGGATGTGGTGTCTGAGATAGTTTGTATCTTAAAGACTGAGGCACA	3180
Qy	3181	GAAGCGAAACGAGAACACACTGTTTGTGAGACACAGTTGTCCAAATGTTTCTGCGCCAGCT	3240
Db	3181	GAAGCGAAACGAGAACACACTGTTTGTGAGACACAGTTGTCCAAATGTTTCTGCGCCAGCT	3240
Qy	3241	CCGCCCCCTTTTGTATGACACTTCTCTCCACCCCTGCACAGCAGTGTGCCCTCATTC	3300
Db	3241	CCGCCCCCTTTTGTATGACACTTCTCTCCACCCCTGCACAGCAGTGTGCCCTCATTC	3300
Qy	3301	TTTTAAATTTTAAAGATGAATGGCAGATGCTAGTAATTCACAGAAATGGCCCTCTGTGGG	3360
Db	3301	TTTTAAATTTTAAAGATGAATGGCAGATGCTAGTAATTCACAGAAATGGCCCTCTGTGGG	3360
Qy	3361	GGTGGGTCTGAGGGAAGTCAAGTATAAAACATTTGCTGGAGTTTGTCAATGGGCTGT	3420
Db	3361	GGTGGGTCTGAGGGAAGTCAAGTATAAAACATTTGCTGGAGTTTGTCAATGGGCTGT	3420
Qy	3421	GCATTTTATATATATGTTTGTAAATGCATGTACAGCCCTTGTTCATGTTTCTCTAAAA	3480
Db	3421	GCATTTTATATATATGTTTGTAAATGCATGTACAGCCCTTGTTCATGTTTCTCTAAAA	3480
Qy	3481	GCAGAAATTTTGAACATTTGTTTGTATAGGAATATTTGTGCGCACTGCTGTGGACTG	3540
Db	3481	GCAGAAATTTTGAACATTTGTTTGTATAGGAATATTTGTGCGCACTGCTGTGGACTG	3540
Qy	3541	TTTTCTTTGGCTAGTACTAGTACCTGTGTGCTTAAACATGAGTTTCAGCCCTTTGGT	3600
Db	3541	TTTTCTTTGGCTAGTACTAGTACCTGTGTGCTTAAACATGAGTTTCAGCCCTTTGGT	3600
Qy	3601	TTTGTAAATACCATGTCAAAATGCAAACTTCAATTCCTCCCATTTAGCTTTATTAAACTG	3660
Db	3601	TTTGTAAATACCATGTCAAAATGCAAACTTCAATTCCTCCCATTTAGCTTTATTAAACTG	3660
Qy	3661	ACGTTCTCTCAAAACCTTCTGCTGAATGGTACTCAGATGTGCAATTCACATACAGATGTG	3720
Db	3661	ACGTTCTCTCAAAACCTTCTGCTGAATGGTACTCAGATGTGCAATTCACATACAGATGTG	3720
Qy	3721	TTTTGAAGTGGGTGCTACCTTCTTACCTAATAGATGTGTAATAGAACTTTTGTAAAGTC	3780
Db	3721	TTTTGAAGTGGGTGCTACCTTCTTACCTAATAGATGTGTAATAGAACTTTTGTAAAGTC	3780
Qy	3781	AAAAAAAAAAAAAAAAAAAA 3798	
Db	3781	AAAAAAAAAAAAAAAAAAAA 3798	
RESULT 2			
AAD21574			
ID	AAD21574 standard; DNA; 3537 BP.		
XX	AC AAD21574;		
XX	AC AAD21574;		
XX	28-JAN-2002 (first entry)		
XX	Human SART3-IC-HER-2 membrane distal intracellular domain fusion DNA.		
XX	Immunostimulatory fusion protein; IFP; antigen component; therapy;		
XX	immunostimulatory component; T-cell mediated immune response; DC;		
XX	dendritic cell; colon cancer; breast carcinoma; ovarian cancer;		
XX	human; HER-2 membrane distal intracellular domain; SART3-IC;		
XX	squamous cell carcinoma antigen; fusion DNA; ds.		
OS	Homo sapiens.		
XX	Key		
PH	Location/Qualifiers		

FT CDS 1..3537  
FT /\*tag= a  
FT /product= "Human SART3-IC-mature HER-2 membrane  
FT distal intracellular domain fusion protein"  
FT /note= "CDS does not include stop codon"  
FT /partial  
XX  
PN WO200174855-A2.  
PD 11-OCT-2001.  
XX  
XX 30-MAR-2001; 2001WO-US10515.  
PF  
XX 30-MAR-2000; 2000US-193504P.  
PR  
XX (DEND-) DENDREON CORP.  
XX  
XX Laus R, Vidovic D, Graddis T;  
XX WPI; 2001-662965/76.  
DR P-PSDB; AAE13123.  
DR  
XX An immunostimulatory fusion protein comprising the intracellular domain  
PT of HER-2 and an antigen elicits an immune response to the antigen and  
PT is useful for the treatment of associated cancer associated -  
XX  
XX Disclosure; Page 58-59; 59pp; English.  
XX  
CC The invention relates to immunostimulatory fusion proteins (IFP) and  
CC nucleic acid molecules encoding such proteins. The IFPs comprise a  
CC polypeptide antigen component and an immunostimulatory component derived  
CC from the intracellular domain of HER-2 protein which is effective to  
CC elicit a protective dendritic cell (DC)-induced T-cell mediated cellular  
CC immune response to the antigen. IFP or superactivated dendritic cells  
CC are used to treat cancer e.g. breast carcinoma, ovarian and colon cancer  
CC associated with a particularly antigen. The present sequence is a  
CC fusion DNA which comprises DNA molecules encoding human squamous cell  
CC carcinoma antigen, SART3-IC and mature human HER-2 membrane distal  
CC intracellular domain.  
XX  
SQ Sequence 3537 BP; 931 A; 885 C; 1048 G; 673 T; 0 other;

Query Match 76.0%; Score 2886; DB 22; Length 3537;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 2886; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 12 ATGGCGACTGCGCGCGAAACCTCGCTTCAGAACCCGAGGCTGAGTCCAAAGCTGGGCC 71  
DB 1 ATGGCGACTGCGCGCGAAACCTCGCTTCAGAACCCGAGGCTGAGTCCAAAGCTGGGCC 60  
QY 72 AAGGCTGACGGAGAGGAGGATGAGTTAAGCGCGCTAGGACAAAGGAGAAAGGTGTTATCG 131  
DB 61 AAGGCTGACGGAGAGGAGGATGAGTTAAGCGCGCTAGGACAAAGGAGAAAGGTGTTATCG 120  
QY 132 CGGGCTGTGGCGCTGCGACATACAAAGCCATGGGGCCAGCGTGGGATCAGCAGAGGAA 191  
DB 121 CGGGCTGTGGCGCTGCGACATACAAAGCCATGGGGCCAGCGTGGGATCAGCAGAGGAA 180  
QY 192 GCGCTGAGCGAGCGATGGGGATGAGTACGCCATGGCTTCCTCCGCGAGAGCTCCGCC 251  
DB 181 GCGCTGAGCGAGCGATGGGGATGAGTACGCCATGGCTTCCTCCGCGAGAGCTCCGCC 240  
QY 252 GGGGAGTACGAGTGGGAATATGACGAAGAGGAGGAGAGAAACACAGCTGGAGATTGAGAGA 311  
DB 241 GGGGAGTACGAGTGGGAATATGACGAAGAGGAGGAGAGAAACACAGCTGGAGATTGAGAGA 300  
QY 312 CTGGAGGACGAGTCTTATCAACGCTATGACTACAACTGCCATGTCGATGATCAGA 371  
DB 301 CTGGAGGACGAGTCTTATCAACGCTATGACTACAACTGCCATGTCGATGATCAGA 360  
QY 372 CTGCTCAGGCTGGAAGGGAGGCTTACCAAGGTGAGGATGGCCCGCAGAAAGATGAGTGA 431  
DB 361 CTGCTCAGGCTGGAAGGGAGGCTTACCAAGGTGAGGATGGCCCGCAGAAAGATGAGTGA 420

QY 432 ATCTTTCCCTTGACTGAAGAGCTCTGGCTGGAGTGGCTGCATGACGAGATCAGCATGGCC 491  
DB 421 ATCTTTCCCTTGACTGAAGAGCTCTGGCTGGAGTGGCTGCATGACGAGATCAGCATGGCC 480  
QY 492 CAGGATGGCTGGACAGAGAGCAGCTGTATGACCTCTTTGAGAAAGCCGTTGAAGATATAC 551  
DB 481 CAGGATGGCTGGACAGAGAGCAGCTGTATGACCTCTTTGAGAAAGCCGTTGAAGATATAC 540  
QY 552 ATTTGCTCTAACATTTGGCTAGAGTATGCCAGTACTCAGTTGGTGGGATTCGTCAGAAA 611  
DB 541 ATTTGCTCTAACATTTGGCTAGAGTATGCCAGTACTCAGTTGGTGGGATTCGTCAGAAA 600  
QY 612 GGTGGCTTTGAGAAAGTTGCTCCGTTGTTGAAAGGGCTCTCTCGCTGCTGTTGTTTACAT 671  
DB 601 GGTGGCTTTGAGAAAGTTGCTCCGTTGTTGAAAGGGCTCTCTCGCTGCTGTTTACAT 660  
QY 672 ATGACCAAAGGACTCGCCCTCTGGGAGGCTTACCAGAGTATGAAAGTGGGATTCGTA 731  
DB 661 ATGACCAAAGGACTCGCCCTCTGGGAGGCTTACCAGAGTATGAAAGTGGGATTCGTA 720  
QY 732 GCTGCTCGCTTTGAGAAAGTCCACAGTCTTTTCCGGCGACAGTTGGGATCCCATCTCTAT 791  
DB 721 GCTGCTCGCTTTGAGAAAGTCCACAGTCTTTTCCGGCGACAGTTGGGATCCCATCTCTAT 780  
QY 792 GATATGGAGGGCCACATTTGCGAGTATGAAGATGCTCAGAAAGTGGTCCAAATACAGAGTCA 851  
DB 781 GATATGGAGGGCCACATTTGCGAGTATGAAGATGCTCAGAAAGTGGTCCAAATACAGAGTCA 840  
QY 852 GTAATTCAGACTATAACAAAGCTACAGAGCTGGGAGAAATATAAACCCCTATGAAGAA 911  
DB 841 GTAATTCAGACTATAACAAAGCTACAGAGCTGGGAGAAATATAAACCCCTATGAAGAA 900  
QY 912 GCACGTGTCAGGCGAGAGGACCAAGGCTGGCAGATATCAGACATATATCATGATTTGAG 971  
DB 901 GCACGTGTCAGGCGAGAGGACCAAGGCTGGCAGATATCAGACATATATCATGATTTGAG 960  
QY 972 ATGAAAATTTGGCGATCCTCGCATTCAGTTGATCTTTGAGCGCGCCCTGGTCGAGAAC 1031  
DB 961 ATGAAAATTTGGCGATCCTCGCATTCAGTTGATCTTTGAGCGCGCCCTGGTCGAGAAC 1020  
QY 1032 TGCCCTTGCCAGACTTATGATTCGTTACAGTACAGTACAGTACAGTACAGTACAGTAC 1091  
DB 1021 TGCCCTTGCCAGACTTATGATTCGTTACAGTACAGTACAGTACAGTACAGTACAGTAC 1080  
QY 1092 AAGGATTTGGTGTATCTGTACATAACCGCTATTAGAAACTGCCCTGGACACTGGCC 1151  
DB 1081 AAGGATTTGGTGTATCTGTACATAACCGCTATTAGAAACTGCCCTGGACACTGGCC 1140  
QY 1152 TTATGAGTTCGGTACCTCTTGCCCATGGAGAGACATGAGTTGATCATCAAGTAATTTCT 1211  
DB 1141 TTATGAGTTCGGTACCTCTTGCCCATGGAGAGACATGAGTTGATCATCAAGTAATTTCT 1200  
QY 1212 GTAACCTTCGAGAAAGCTTTGAAATCCCGCTTCATCCAGGCCACTGATATGTCGAGATT 1271  
DB 1201 GTAACCTTCGAGAAAGCTTTGAAATCCCGCTTCATCCAGGCCACTGATATGTCGAGATT 1260  
QY 1272 TGGCAGGACATACCTTTGATTACCTGAGGAGAAAGGTTGATTTCAAAACAAGACTCCAGTAAA 1331  
DB 1261 TGGCAGGACATACCTTTGATTACCTGAGGAGAAAGGTTGATTTCAAAACAAGACTCCAGTAAA 1320  
QY 1332 GAGCTGGAGGAGTTGAGGGCCGCCCTTTACTCGTGCCTTTGGAGTATCTCAAGCAGAGGTG 1391  
DB 1321 GAGCTGGAGGAGTTGAGGGCCGCCCTTTACTCGTGCCTTTGGAGTATCTCAAGCAGAGGTG 1380  
QY 1392 GAAGAGCCTTTCAATGAGAGTGGTGCATCAAGCTCCGTCGTTATTCAGAACTGGGCTAGG 1451  
DB 1381 GAAGAGCCTTTCAATGAGAGTGGTGCATCAAGCTCCGTCGTTATTCAGAACTGGGCTAGG 1440  
QY 1452 ATTGAGGCTCGACTGTGCAATAAATGAGAAAGCTCGGAACTCTGGGATAGCATCATG 1511  
DB 1441 ATTGAGGCTCGACTGTGCAATAAATGAGAAAGCTCGGAACTCTGGGATAGCATCATG 1500

QY	1512	ACCAGAGAAATGCCAAGTACGCCAACATGTGGCTAGAGTATTACAACTGGGAAGAGCT	1571		
Db	1501	ACCAGAGAAATGCCAAGTACGCCAACATGTGGCTAGAGTATTACAACTGGGAAGAGCT	1560		
QY	1572	CATGGTACACCCAGCACTGCCGGAAGGCTCTGCACCGGGCCGTCCAGTGCACAGTGAC	1631		
Db	1561	CATGGTACACCCAGCACTGCCGGAAGGCTCTGCACCGGGCCGTCCAGTGCACAGTGAC	1620		
QY	1632	TACCACAGACGCTCTGCGAAAGTGTACTACCAATGGAGAGACAGAAAGTCTTTAGAA	1691		
Db	1621	TACCACAGACGCTCTGCGAAAGTGTACTACCAATGGAGAGACAGAAAGTCTTTAGAA	1680		
QY	1692	GATTGGGATATAGCTGTTTCAAAAACCTGAAACCCGATTAGCTCGTGTCAATGACAGAGA	1751		
Db	1681	GATTGGGATATAGCTGTTTCAAAAACCTGAAACCCGATTAGCTCGTGTCAATGACAGAGA	1740		
QY	1752	ATGAAGGCTGCAGAGAAGGAGCAGCCCTTGTGCAGCAAGAAAGAAAGGCTGAACAA	1811		
Db	1741	ATGAAGGCTGCAGAGAAGGAGCAGCCCTTGTGCAGCAAGAAAGAAAGGCTGAACAA	1800		
QY	1812	CGGAAAGAGCTCGGGCTGAGAAGAAAGCGTTTAAAAAGAAAGAAAGATCAGAGGCCA	1871		
Db	1801	CGGAAAGAGCTCGGGCTGAGAAGAAAGCGTTTAAAAAGAAAGAAAGATCAGAGGCCA	1860		
QY	1872	GAGAAAGCGGAGCAGATGAGGACGATGAGAAAGATGGGCGATGATGAAGAAGAGCAG	1931		
Db	1861	GAGAAAGCGGAGCAGATGAGGACGATGAGAAAGATGGGCGATGATGAAGAAGAGCAG	1920		
QY	1932	CCTTCCAAACCCAGAAAGGTCGAGNACAGCATCCCTGCAGCTGGAGAAACACAAATGTA	1991		
Db	1921	CCTTCCAAACCCAGAAAGGTCGAGNACAGCATCCCTGCAGCTGGAGAAACACAAATGTA	1980		
QY	1992	GAACTAGCAGCAGGCGCGCTGGGAAATGTCTGCGCTAGATGTGGAGCCCTTCGGAAG	2051		
Db	1981	GAACTAGCAGCAGGCGCGCTGGGAAATGTCTGCGCTAGATGTGGAGCCCTTCGGAAG	2040		
QY	2052	CAGAAGGAGAGGCGCCTCCCTGAAGAGGACATGCCAAAGGTGCTGCAGCAGACAGCAGC	2111		
Db	2041	CAGAAGGAGAGGCGCCTCCCTGAAGAGGACATGCCAAAGGTGCTGCAGCAGACAGCAGC	2100		
QY	2112	AAGGACACATCAGCGTCTTTGTCTAGCAACCTGCCCTACAGATGCAGGAGCGCGACAGC	2171		
Db	2101	AAGGACACATCAGCGTCTTTGTCTAGCAACCTGCCCTACAGATGCAGGAGCGCGACAGC	2160		
QY	2172	AAGCTCAGGCGCCTCTGAGGCGCTGTGGGAGGTGCTCCAGATCCGACCCATCTTCAGC	2231		
Db	2161	AAGCTCAGGCGCCTCTGAGGCGCTGTGGGAGGTGCTCCAGATCCGACCCATCTTCAGC	2220		
QY	2232	AACCGTGGGGATTTCCGAGGTTTACTGCTACGTGGAGTTTAAAGAGAGAAATCAGCCCTT	2291		
Db	2221	AACCGTGGGGATTTCCGAGGTTTACTGCTACGTGGAGTTTAAAGAGAGAAATCAGCCCTT	2280		
QY	2292	CAGGCACTGGAGATGGACCGGAAAGTGTAGAAGGAGGCGCAATGTTTGTCCCCCTGT	2351		
Db	2281	CAGGCACTGGAGATGGACCGGAAAGTGTAGAAGGAGGCGCAATGTTTGTCCCCCTGT	2340		
QY	2352	GTGGATAAGAGCAAAACCCGATTTTAAAGTGTTCAGGTACAGCACTTCCTTAGAGAAA	2411		
Db	2341	GTGGATAAGAGCAAAACCCGATTTTAAAGTGTTCAGGTACAGCACTTCCTTAGAGAAA	2400		
QY	2412	CACAAGCTGTTTATCTTCAGGCGCTGCTTCTCTCTACTAAAGAGAACTAGAAGAAATC	2471		
Db	2401	CACAAGCTGTTTATCTTCAGGCGCTGCTTCTCTCTACTAAAGAGAACTAGAAGAAATC	2460		
QY	2472	TGTAAGGCTCATGSCACCGTGAAGACCTCAGGCTGGTCCCAACCGGGCTGGCAACCA	2531		
Db	2461	TGTAAGGCTCATGSCACCGTGAAGACCTCAGGCTGGTCCCAACCGGGCTGGCAACCA	2520		
QY	2532	AAGGCGCTGGCTACGTGAGGTATGAAATGAATCCAGGGGTCCAGGCTGTGTATGAAG	2591		
Db	2521	AAGGCGCTGGCTACGTGAGGTATGAAATGAATCCAGGGGTCCAGGCTGTGTATGAAG	2580		
QY	2592	ATGGACGGCATGACTATCAAGAGAAACATCATCAAAAGTGGCAATCAGCAACCCCTCCTCAG	2651		
Db	2581	ATGGACGGCATGACTATCAAGAGAAACATCATCAAAAGTGGCAATCAGCAACCCCTCCTCAG	2640		
QY	2652	AGGAAAGTTCAGAGAAGCCAGAGACCAAGGAGGACCAAGTGGCCCATGCTTTTGGCG	2711		
Db	2641	AGGAAAGTTCAGAGAAGCCAGAGACCAAGGAGGACCAAGTGGCCCATGCTTTTGGCG	2700		
QY	2712	CAGACATACGGAGCGAGGGGGAAGGAAGGAGCGAGCTGTCTTACTGCTCGTCCCTCG	2771		
Db	2701	CAGACATACGGAGCGAGGGGGAAGGAAGGAGCGAGCTGTCTTACTGCTCGTCCCTCG	2760		
QY	2772	CAGCGCCCAAGTGTCTGACGCTCCTCAGGCTGAGAAGCGGCCCTGCGCGGCTCCTGAGTT	2831		
Db	2761	CAGCGCCCAAGTGTCTGACGCTCCTCAGGCTGAGAAGCGGCCCTGCGCGGCTCCTGAGTT	2820		
QY	2832	GCCGCCCCAGCAGCCACCGAGGACCCCAAGATGTCCAATGCCGATTTTGCCCAAGCTGTTT	2891		
Db	2821	GCCGCCCCAGCAGCCACCGAGGACCCCAAGATGTCCAATGCCGATTTTGCCCAAGCTGTTT	2880		
QY	2892	CTGAGA 2897			
Db	2881	CTGAGA 2886			
RESULT 3					
AAAF16121					
ID	AAAF16121 standard; cdna; 3355 BP.				
XX	AAAF16121;				
AC	AAAF16121;				
DT	13-MAR-2001 (first entry)				
XX	Human prostate cancer antigen nucleotide sequence SEQ ID NO:556.				
DE	Human; prostate cancer; prostate cancer antigen; detection; diagnosis;				
XX	neuroprotective; cytosolic; cardioactive; immunomodulatory; muscular;				
KW	vulnary; gastrointestinal; nephrotropic; antiinfective; gynaecological;				
KW	antibacterial; gene therapy; neural; immune; reproductive; renal;				
KW	gastrointestinal; pulmonary; cardiovascular; proliferative disorder;				
KW	wound; infectious disease; ss.				
XX	Homo sapiens.				
OS	Homo sapiens.				
PN	WO200055174-A1.				
XX	21-SEP-2000.				
PD	21-SEP-2000.				
XX	08-MAR-2000; 2000WO-US05988.				
PF	12-MAR-1999; 99US-0124270.				
XX	(HUMA-) HUMAN GENOME SCI INC.				
XX	(ROSE/) ROSEN C A.				
PA	Rosen CA, Ruben SM;				
PI	WPI: 2000-587513/55.				
XX	P-PSDB; AAB56918.				
DR	Prostate cancer associated gene sequences, referred to as prostate				
PT	cancer antigens, useful for treatment, prevention, and diagnosis of				
PT	disorders such as prostate cancer -				
XX	Claim 1; Page 1014-1015; 2338pp; English.				
PS	AAAF15566 to AAFA15505 encode the human prostate cancer associated				
XX	proteins, called prostate cancer antigens, given in AAB56363 to AAB57302.				
CC	The prostate cancer antigens can have neuroprotective, cytosolic,				
CC	cardioactive, immunomodulatory, muscular, vulnary, gastrointestinal,				
CC	neurotropic, antiinfective, gynaecological and antibacterial activities,				
CC	and can be used in gene therapy. The prostate cancer antigen				
CC	polynucleotides may be used for detection of prostate cancer, chromosome				
CC	identification, as chromosome markers, and for numerous other diagnostic				

or research purposes. The prostate cancer antigens may be used to treat disorders such as neural, immune, muscular, reproductive, gastrointestinal, pulmonary, cardiovascular, renal, and proliferative disorders, wounds, and infectious diseases. AAF16506 to AAF16514 to AAF57303 represent sequences used in the exemplification of the present invention.

Sequence 3355 BP; 923 A; 716 C; 853 G; 862 T; 1 other;

Query Match	67.5%;	Score 2563.2;	DB 21;	Length 3355;
Best Local Similarity	99.8%;	Pred. No. 0;		
Matches 2576;	Conservative	0;	Mismatches 3;	Indels 1; Gaps 1;
QY	1205	AATTTCTGAACCTTCGAGAAAGCTTTGAATGCGGCTTCATCCAGGCCACTGATTATGT	1264	
DB	386	AACCTCCAGTAACCTTCGAGAAAGCTTTGAATGCGGCTTCATCCAGGCCACTGATTATGT	445	
QY	1265	GGAGATTGGCAGGCATACCTTGATTACCTGAGGAGAAGGTTGATTTCAACAAGACTC	1324	
DB	446	GGAGATTGGCAGGCATACCTTGATTACCTGAGGAGAAGGTTGATTTCAACAAGACTC	505	
QY	1325	CAGTAAGAGCTGGAGAGTTGAGGCGGCTTTACTCGCTTGGAGTATCTGAAGCA	1384	
DB	506	CAGTAAGAGCTGGAGAGTTGAGGCGGCTTTACTCGCTTGGAGTATCTGAAGCA	565	
QY	1385	GGAGTGGAGAGCGTTTCAATGAGAGTGGTGATCCAGCTGCGTGATTATGCAAGACTG	1444	
DB	566	GGAGTGGAGAGCGTTTCAATGAGAGTGGTGATCCAGCTGCGTGATTATGCAAGACTG	625	
QY	1445	GGCTAGATTGAGGCTCGACTGTGCAATACATGCAAGAGCTCGGGAACCTCGGATAG	1504	
DB	626	GGCTAGATTGAGGCTCGACTGTGCAATACATGCAAGAGCTCGGGAACCTCGGATAG	685	
QY	1505	CATCATGACCAGAGAAATGCCAAGTACGCCAACATGTGGCTAGAGTATTACAACCTGGA	1564	
DB	686	CATCATGACCAGAGAAATGCCAAGTACGCCAACATGTGGCTAGAGTATTACAACCTGGA	745	
QY	1565	AGAGCTCATGTTGACACCCAGCAGCTGCGGGAAGGCTCTGCACCGGGCGTCCAGTGCAC	1624	
DB	746	AGAGCTCATGTTGACACCCAGCAGCTGCGGGAAGGCTCTGCACCGGGCGTCCAGTGCAC	805	
QY	1625	CAGTACTACCCAGACAGCTCTGCGAAGTGTACTCACCTGAGGAGCAGACAGAGTTTC	1684	
DB	806	CAGTACTACCCAGACAGCTCTGCGAAGTGTACTCACCTGAGGAGCAGACAGAGTTTC	865	
QY	1685	TTTAGAAGATTGGGATATAGCTGTTTCAGAAAACCTGAAACCCGATTAGCTCGTCAATGA	1744	
DB	866	TTTAGAAGATTGGGATATAGCTGTTTCAGAAAACCTGAAACCCGATTAGCTCGTCAATGA	925	
QY	1745	GCAGAGAATGAAGGCTGCAGAGAAGGAGCAGCCCTTGTGCAGCAAGAAAGAAAGGC	1804	
DB	926	GCAGAGAATGAAGGCTGCAGAGAAGGAGCAGCCCTTGTGCAGCAAGAAAGAAAGGC	985	
QY	1805	TGAACAACGGAAGAGCTCGGCTTCAGAGAAGGCTTTAAAGAAAGAAAGAAAGATCAG	1864	
DB	986	TGAACAACGGAAGAGCTCGGCTTCAGAGAAGGCTTTAAAGAAAGAAAGAAAGATCAG	1045	
QY	1865	AGGCCAGAGAAGCGGAGCAGATGAGGACGATGAGAAAGAGTGGGGCGATGATGAAGA	1924	
DB	1046	AGGCCAGAGAAGCGGAGCAGATGAGGACGATGAGAAAGAGTGGGGCGATGATGAAGA	1105	
QY	1925	AGAGCAGCCTTCAACCGCAGAGGTCGAGAACAGCATCCCTGCAGCTGGAGAGAACACA	1984	
DB	1106	AGAGCAGCCTTCAACCGCAGAGGTCGAGAACAGCATCCCTGCAGCTGGAGAGAACACA	1165	
QY	1985	AAATGTAGAGTAGCAGAGGCGGCTCGGAAATGCTGCTGCTAGATGTGGAGCCCC	2044	
DB	1166	AAATGTAGAGTAGCAGAGGCGGCTCGGAAATGCTGCTGCTAGATGTGGAGCCCC	1225	
QY	2045	TTTCAAGCAGAGAAGGAGGAGGCTTCCCTGGAAGAGGAGACATGCCAAGGCTGTCACGA	2104	
DB	1226	TTTCAAGCAGAGAAGGAGGAGGCTTCCCTGGAAGAGGAGACATGCCAAGGCTGTCACGA	1285	

QY	2105	CAGACGACAGGACAGCATCACCGCTTTTGTACGAAACCTGCCCTACAGCATGCGAGGCC	2164	
DB	1286	CAGACGACAGGACAGCATCACCGCTTTTGTACGAAACCTGCCCTACAGCATGCGAGGCC	1345	
QY	2165	GGACACGAGCTCAGGCCACTCTTCAGAGCCTGTGGGAGGTGGTCCAGATCCGACCCAT	2224	
DB	1346	GGACACGAGCTCAGGCCACTCTTCAGAGCCTGTGGGAGGTGGTCCAGATCCGACCCAT	1405	
QY	2225	CTTCACGAAACCGTGGGGATTTCGAGGTTACTCTACGTGGAGTTTAAAGAGAGAAATC	2284	
DB	1406	CTTCACGAAACCGTGGGGATTTCGAGGTTACTCTACGTGGAGTTTAAAGAGAGAAATC	1465	
QY	2285	AGCCCTTCAGGCACCTGGAGATGGACCGGAAAGTGTAGAAGGAGAGCCCAATGTTTCTTC	2344	
DB	1466	AGCCCTTCAGGCACCTGGAGATGGACCGGAAAGTGTAGAAGGAGAGCCCAATGTTTCTTC	1525	
QY	2345	CCCCTGTGTGATAGAGCAAAACCCCGATTTTAAGGTGTTTCAGGTACAGCACTTCCCT	2404	
DB	1526	CCCCTGTGTGATAGAGCAAAACCCCGATTTTAAGGTGTTTCAGGTACAGCACTTCCCT	1585	
QY	2405	AGAGAAACCAAGCTTTCATCTCAGGCCTGCTCTCTCTGTACTTAAAGAGGAACCTAGA	2464	
DB	1586	AGAGAAACCAAGCTTTCATCTCAGGCCTGCTCTCTCTGTACTTAAAGAGGAACCTAGA	1645	
QY	2465	AGAAATCTGTAGGCTCATGGCACCTGAAGACCTCAGGCTGGTCCACCAACCGGCTGG	2524	
DB	1646	AGAAATCTGTAGGCTCATGGCACCTGAAGACCTCAGGCTGGTCCACCAACCGGCTGG	1705	
QY	2525	CAACCAAGGCGCTGGCCTAGCTGAGTATGAAATGAATCCAGGCTGCGAGGCTGT	2584	
DB	1706	CAACCAAGGCGCTGGCCTAGCTGAGTATGAAATGAATCCAGGCTGCGAGGCTGT	1765	
QY	2585	GATGAAGTGGAGCGGATGACTATCAAGAGAAACATCATCAAGTGGCAATCAGCAACCC	2644	
DB	1766	GATGAAGTGGAGCGGATGACTATCAAGAGAAACATCATCAAGTGGCAATCAGCAACCC	1825	
QY	2645	TCCTCAGAGAAAGTTCCAGAGAACCCAGAGACAGGCAAGGACCAAGTGGGCCCATGCT	2704	
DB	1826	TCCTCAGAGAAAGTTCCAGAGAACCCAGAGACAGGCAAGGACCAAGTGGGCCCATGCT	1885	
QY	2705	TTTGCCGACAGCATACGGAGCGAGGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG	2764	
DB	1886	TTTGCCGACAGCATACGGAGCGAGGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG	1945	
QY	2765	TGCGCTGACGCGCCCAAGTGTCTCAGGCTGAGAACGGCCCTGCCCGGCTCC	2824	
DB	1946	TGCGCTGACGCGCCCAAGTGTCTCAGGCTGAGAACGGCCCTGCCCGGCTCC	2005	
QY	2825	TGCGCTGCGCCCGCAGCAGCCAGGACCCAGAGATGTCCTCAATGCCGATTTTGCCAA	2884	
DB	2006	TGCGCTGCGCCCGCAGCAGCCAGGACCCAGAGATGTCCTCAATGCCGATTTTGCCAA	2065	
QY	2885	GCTGTTCTGCAAAAGTGAACGGAGCGCTGGGAGACAGAAATGCTTACTTCACTCTGG	2944	
DB	2066	GCTGTTCTGCAAAAGTGAACGGAGCGCTGGGAGACAGAAATGCTTACTTCACTCTGG	2125	
QY	2945	CCCGGGGACCTCCCAACCCAGCAGTGCATCTGGGGATGGACAGGCTTGGTGTCTGG	3004	
DB	2126	CCCGGGGACCTCCCAACCCAGCAGTGCATCTGGGGATGGACAGGCTTGGTGTCTGG	2185	
QY	3005	TGCTCGAACACAGATGGCTCCTCGGCTTTAGACAGAAAGGAGGAGGTTCTAAGTCA	3064	
DB	2186	TGCTCGAACACAGATGGCTCCTCGGCTTTAGACAGAAAGGAGGAGGTTCTAAGTCA	2245	
QY	3065	AGAGCCTTTCAGTGTCTCCTCATATTGAGGCGAGTGGCAAGAGTGACCTCTGCAGG	3124	
DB	2246	AGAGCCTTTCAGTGTCTCCTCATATTGAGGCGAGTGGCAAGAGTGACCTCTGCAGG	2305	
QY	3125	CTGGGCGGAGGATGGTGTCTCTGAGATAGTTTGTATCTTAAAGACTGAGCACAGAG	3184	
DB	2306	CTGGGCGGAGGATGGTGTCTCTGAGATAGTTTGTATCTTAAAGACTGAGCACAGAG	2365	
QY	3185	CGAAACGAGACACACTGTTTTTTGAGACACACTGTTTCCAAATGTTTCTGGGCCAGCTCGG	3244	

















QY	393	CTTACCAAGCTGAGGATGGCCCGCCAGAAATCAGTGAATCTTTCCCTTGACTGAAGAG	452
Db	141	CTTACCAAGCTGAGGATGGCCCGCCAGAAATCAGTGAATCTTTCCCTTGACTGAAGAG	200
QY	453	CTCTGGCTGGAGTGGCTGCGATGAGGAGATCAGCATGCCCGAGGATGGCTGGACACAGAG	512
Db	201	CTCTGGCTGGAGTGGCTGCGATGAGGAGATCAGCATGCCCGAGGATGGCTGGACACAGAG	260
QY	513	CACGTGTATGACCTCTTTTGAGAAAGCCGTGAAGGATTACATTTGTCTTCAACATTTGGCTA	572
Db	261	CACGTGTATGACCTCTTTTGAGAAAGCGTGAAGGATTACATTTGTCTTCAACATTTGGCTA	320
QY	573	GAGTATGCCAGTACTCAGTTGGTGGATTGGTTCAGAAAGTGGCCCTTGAGAAAGTTGCG	632
Db	321	GAGTATGCCAGTACTCAGTTGGTGGATTGGTTCAGAAAGTGGCCCTTGAGAAAGTTGCG	380
QY	633	TCGCTGTTTGAAGGGCTCTCTCGCTCTGTTGGTTTACATATGACCAAGGACTCG	687
Db	381	TCGCTGTTTGAAGGGCTCTCTCGCTCTGTTGGTTTACATATGACCAAGGACTCG	435

CC	with (III) or (II). (III) is useful in design and preparation of
CC	ribozyme molecules for inhibiting expression of the tumour polypeptides
CC	and proteins in tumour cells; and to isolate a full length gene from a
CC	suitable library e.g., a tumour cDNA library using well known
CC	techniques.
XX	
SQ	Sequence 432 BP; 108 A; 86 C; 105 G; 132 T; 1 other;
	Query Match 10.4%; Score 393.4; DB 24; Length 432;
	Best Local Similarity 98.8%; Pred. No. 2.le-96;
	Matches 427; Conservative 0; Mismatches 2; Indels 3; Gaps 3;
QY	3063 CAAGAGCCTTTCAGTGCTCCCTCATATGAGGGCACTGGCAGAAAAGTGACCACACTCTGCA 3122
Db	2 CAAGAGCCTTTCAGTGCTCCCTCATATGAGGGCACTGGCAGAAAAGTGACCACACTCTGCA 61
QY	3123 GGCTGGCGGCAGGATGTGGTGCTCGAGATAGTTTGTATCTTAAAGACTGAGGCACAGA 3182
Db	62 GGCTGGG-CCAGGATGTGGTGCTCGAGATAGTTTGTATCTTAAAGACTGAGGCACAGA 120
QY	3183 AGCGAAACGAGAACACA-CTGTTTTTTTGAGACACAGTTGTCCAAATGTTTTCTGGCCAGCTC 3241
Db	121 AGCGAAACGAGAACACACCTGTTTTTGAGACACAGTTGTCCAAATGTTTTCTGGCCAGCTC 180
QY	3242 CGGCCCTTTTTGTATGACACTTCTCTTCCACCTGCACAGCACATGTGCCCGTCAATTCT 3301
Db	181 CGGCCCTTTTTGTATGACACTTCTCTTCCACCTGCACAGCACATGTGCCCGTCAATTCT 240
QY	3302 TTTTAATTTTAAAGATGAATGGCAGATGCTAGTAGTAATTCACAGAAATGGCCTCTTGTGGGG 3361
Db	241 TTTTAATTTTAAAGATGAATGGCAGATGCTAGTAGTAATTCACAGAAATGGCCTCTTGTGGGG 300
QY	3362 GTGGGCTGAGGGAAGTCAGCTATAAACAATTTGCTGGAGTTTGTTCATATGGGCTGTG 3421
Db	301 GTGGGCTGAGGGAAGTCAGCTATAAACAATTTGCTGGAGTTTGTTCATATGGGCTGTG 360
QY	3422 CATTTTT-APATATTCGTTGTTGFAATGACATGTCAGCCCTGTTTTCATGTTTCCATAAAA 3480
Db	361 CATTTTTAATATTCGTTGTTGFAATGACATGTCAGCCATTCATGTTTCCATAAAA 420
QY	3481 GCAGATAATTG 3492
Db	421 GCAGATAATTG 432
	RESULT 10
	ABL79050
ID	ABL79050 standard; cdNA; 430 BP.
XX	
XX	ABL79050:
XX	
DT	17-MAY-2002 (first entry)
XX	
DE	Human ovarian cancer related cdNA clone SEQ ID NO:2028.
XX	
KW	Human; ovarian cancer; ovarian tumour; cytostatic; gene; ss.
XX	
OS	Homo sapiens.
XX	
PN	WO200192581-A2.
XX	
PD	06-DEC-2001.
XX	
Pf	29-MAY-2001; 2001WO-US17756.
XX	
PR	26-MAY-2000; 2000US-207484P.
XX	
PA	(CORI-) CORIXA CORP.
XX	
PI	Algate PA, Harlocker SL, Jones R;
XX	
DR	WPI; 2002-122075/16.
XX	

PT Composition for therapy and diagnosis of ovarian cancer comprising  
PT polypeptide of an ovarian tumor polypeptide, polynucleotide encoding  
PT polypeptide, antibody specific to polypeptide or T cell expressing  
PT polypeptide

PS Claim 1; SEQ ID 2028; 489pp; English.

XX The present invention describes a composition (I) comprising: carriers  
CC and immunostimulants; and a polypeptide (II) of an ovarian tumor  
CC polypeptide encoded by a polynucleotide (III) having a cDNA sequence  
CC (S1) from the 10912 nucleotide sequences as given in ABL77023 to  
CC ABL87934, (III) encoding (II) having a sequence (S2), a T cell  
CC population of (II), or antigen presenting cells that express (II).  
CC (I) has cytostatic activity. An oligonucleotide (IV) that hybridises to  
CC (S1) can be used for detecting ovarian cancer in a patient's biological  
CC sample preferably serum or ovarian tissue. The method comprises  
CC contacting a biological sample from a patient with (IV), detecting the  
CC amount of polynucleotide hybridising to (IV) and comparing the amount to  
CC a predetermined cutoff value and thereby detecting ovarian cancer in the  
CC patient, where the amount of polynucleotide hybridising to (IV) is  
CC detected preferably by polymerase chain reaction (PCR). (I) comprising  
CC (III) and/or (II) is useful for stimulating and/or expanding T cells  
CC specific for an ovarian tumor protein comprising contacting T cells  
CC with (III) or (II). (III) is useful in design and preparation of  
CC ribozyme molecules for inhibiting expression of the tumor polypeptides  
CC and proteins in tumor cells; and to isolate a full length gene from a  
CC suitable library e.g., a tumor cDNA library using well known  
CC techniques.

XX Sequence 430 BP; 108 A; 85 C; 104 G; 132 T; 1 other;

Query Match 10.3%; Score 392.4; DB 24; Length 430;

Best Local Similarity 98.8%; Pred. No. 4e-96;

Matches 426; Conservative 0; Mismatches 2; Indels 3; Gaps 3;

QY 3064 AAGAGCCTTTCAGTCTCCTCATATTGAGGCGAGTGGCAGAAAGTGACCACTCTGCAG 3123

DB 1 AAGAGCCTTTCAGTCTCCTCATATTGAGGCGAGTGGCAGAAAGTGACCACTCTGCAG 60

QY 3124 GCTGGGCCAGGATGGTGTCTCTGAGATAGTATTTGTATCTTAAAGACTGAGGCACAGAA 3183

DB 61 GCTGGG-CCAGATGGTGTCTCTGAGATAGTATTTGTATCTTAAAGACTGAGGCACAGAA 119

QY 3184 GCGAAGCAGAGACACA-CGTGTTTTGAGACAGATGTGCCAAATGTTCTGCCAGCTCC 3242

DB 120 GCGAAGCAGAGACACACCTGTTTTGAGACAGATGTGCCAAATGTTCTGCCAGCTCC 179

QY 3243 GGGCCCTTTTCTATCAGACTTCTCTCCACCCTGCACAGCACATGTGCCCGCTCATCTT 3302

DB 180 GGGCCCTTTTCTATGACACTTCTCTCCACCCTGCACAGCACATGTGCCCGCTCATCTT 239

QY 3303 TTAATTTTAAAAGATGAATGSCAGATGCTAGTAAATTCACAGAAATGGCCTCTTGTGGGG 3362

DB 240 TTAATTTTAAAAGATGAATGSCAGATGCTAGTAAATTCACAGAAATGGCCTCTTGTGGGG 299

QY 3363 TGGGTCTGAGGAAGTACAGCTATAAACAATTTGCTGGAGTTTTGTTCAATGGGCTGTGC 3422

DB 300 TGGGTCTGAGGAAGTACAGCTATAAACAATTTGCTGGAGTTTTGTTCAATGGGCTGTGC 359

QY 3423 ATTTT-ATATTATGTTTGTAAATGACATGTCAGCCCTTCTTCATGTTTCCCTAAAG 3481

DB 360 ATTTTAAATATTATGTTTGTAAATGACATGTCAGCCCTTCTTCATGTTTCCCTAAAG 419

QY 3482 CAGAATATTG 3492

DB 420 CAGAATATTG 430

RESULT 11

AAH98082

ID AAH98082 standard; DNA; 554 BP.

XX

AC AAH98082;

XX

DT

XX

DE

XX

KW

XX

KW

XX

OS

XX

PN

XX

XX

PD

XX

PF

XX

PR

XX

PA

XX

PA

XX

PI

XX

XX

DR

XX

PT

XX

PT

XX

PT

XX

PS

XX

XX

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

```
Db 360 TGCAGGCAGANGCCCTCGGNGCGGAATACCAAGCTTACATCGANTTCGAGATGAAA 419
Qy 979 TTGCGGATCCTGCTCGCATTCAGTTGATCTTTGAGCGGCCCTGGTCGAGACTGCGTTG 1038
Db 420 TCGGGGAT-CTGGCCGATTACGTTGATCTTTGAGCGTTCTGGTGGAGAACTTGCCTGG 478
Qy 1039 TCCAGACTTATGATCCGTTTACAGTACCTAGTACCTAGTACGACAACTGAAAGTAAAGGAT 1098
Db 479 NTCCAGNCTTATGATC-GCCACAGTNGTACCTAGATCCACAGCTGAAGTTAAGGCT 536
Qy 1099 TGGTTT 1104
Db 537 TGGTTT 542

RESULT 12
AAK75790/c
ID AAK75790 standard; DNA; 543 BP.
XX AC
XX AC
XX AC
XX DT
XX 07-NOV-2001 (first entry)
DE Human immune/haematopoietic antigen genomic sequence SEQ ID NO:30602.
XX Human; immune; haematopoietic; immune/haematopoietic antigen; cancer;
KW cytostatic; gene therapy; vaccine; metastasis; ds.
XX Homo sapiens.
OS
XX
XX WO200157182-A2.
XX
XX 09-AUG-2001.
XX
XX 17-JAN-2001; 2001WO-US01354.
XX
XX 31-JAN-2000; 2000US-0179065.
PR 04-FEB-2000; 2000US-0180628.
PR 24-FEB-2000; 2000US-0184664.
PR 02-MAR-2000; 2000US-0186350.
PR 16-MAR-2000; 2000US-0189874.
PR 17-MAR-2000; 2000US-0190076.
PR 18-APR-2000; 2000US-0198123.
PR 19-MAY-2000; 2000US-0205515.
PR 07-JUN-2000; 2000US-0209467.
PR 28-JUN-2000; 2000US-0214886.
PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216647.
PR 07-JUL-2000; 2000US-0216880.
PR 11-JUL-2000; 2000US-0217487.
PR 11-JUL-2000; 2000US-0217496.
PR 14-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220963.
PR 26-JUL-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226868.
PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 12-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 17-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
```





CC from WIPO at ftp.wipo.int/pub/published\_pct\_sequences.

XX Sequence 312 BP; 41 A; 125 C; 66 G; 80 T; 0 other;

Query Match 8.2%; Score 310.4; DB 22; Length 312;  
Best Local Similarity 99.7%; Pred. No. 6.5e-74;  
Matches 311; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 12 ATGGCGACTGCGCGCGGAAACCTCGGCTTCAGAACCCGAGGCTGAGTCCAAAGGCTGGGCC 71  
Dy 312 ATGGCGACTGCGCGCGGAAACCTCGGCTTCAGAACCCGAGGCTGAGTCCAAAGGCTGGGCC 253  
Qy 72 AAGGCTGACGAGAGGAGGATGAGTTAAGCGGCTAGGACAAAGAGAAAGTTTATCG 131  
Dy 252 AAGGCTGACGAGAGGAGGATGAGTTAAGCGGCTAGGACAAAGAGAAAGTTTATCG 193  
Qy 132 CGGCTGTGGCGGCTGCGACATACAAAGACCATGGGCGGCGGATCGGATCAGCAGAGGAA 191  
Dy 192 CGGCTGTGGCGGCTGCGACATACAAAGACCATGGGCGGCGGATCGGATCAGCAGAGGAA 133  
Qy 192 GGCCTGACGAGAGCGATGGGATGAGTACGCCATGGCTTCCTCCGCGGAGAGCTCCGCC 251  
Dy 132 GGCCTGACGAGAGCGATGGGATGAGTACGCCATGGCTTCCTCCGCGGAGAGCTCCGCC 73  
Qy 252 GGGGAGTACGAGTGGGAATATGACGAAGAGAGAGAGAAACCAGCTGGAGATTGAGAGA 311  
Dy 72 GGGGAGTACGAGTGGGAATATGACGAAGAGAGAGAGAAACCAGCTGGAGATTGAGAGA 13  
Qy 312 CTGGAGGAGCAG 323  
Dy 12 CTGGAGGAGCAG 1

## RESULT 14

ABA67045/C  
ID ABA67045 standard; DNA; 312 BP.

XX ABA67045;

XX 01-FEB-2002 (first entry)

XX Human foetal liver single exon nucleic acid probe #15350.

XX Human; foetal liver; gene expression; single exon nucleic acid probe; ss.

XX Homo sapiens.

XX WO200157277-A2.

XX 09-AUG-2001.

XX 30-JAN-2001; 2001WO-US00669.

XX 04-FEB-2000; 2000US-0180312.

XX 26-MAY-2000; 2000US-0207456.

XX 30-JUN-2000; 2000US-0608408.

XX 03-AUG-2000; 2000US-0632366.

XX 21-SEP-2000; 2000US-0234687.

XX 27-SEP-2000; 2000US-0236359.

XX 04-OCT-2000; 2000GB-0024263.

XX (MOLE-) MOLECULAR DYNAMICS INC.

XX Penn SG, Hanzel DK, Chen W, Rank DR;

XX WPI; 2001-483447/52.

XX Human genome-derived single exon nucleic acid probes useful for  
XX analyzing gene expression in human foetal liver.

XX Claim 4; SEQ ID NO 15350; 639pp + sequence listing; English.

XX The invention relates to a single exon nucleic acid probe for

CC measuring human gene expression in a sample derived from human foetal  
CC liver. The single exon nucleic acid probes may be used for predicting,  
CC measuring and displaying gene expression in samples derived from human  
CC foetal liver. The present sequence is a single exon nucleic acid  
CC probe of the invention.  
CC Note: The sequence data for this patent did not form part of the  
CC printed specification, but was obtained in electronic format directly  
CC from WIPO at ftp.wipo.int/pub/published\_pct\_sequences.

XX Sequence 312 BP; 41 A; 125 C; 66 G; 80 T; 0 other;

Query Match 8.2%; Score 310.4; DB 22; Length 312;  
Best Local Similarity 99.7%; Pred. No. 6.5e-74;  
Matches 311; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 12 ATGGCGACTGCGCGCGGAAACCTCGGCTTCAGAACCCGAGGCTGAGTCCAAAGGCTGGGCC 71  
Dy 312 ATGGCGACTGCGCGCGGAAACCTCGGCTTCAGAACCCGAGGCTGAGTCCAAAGGCTGGGCC 253  
Qy 72 AAGGCTGACGAGAGGAGGATGAGTTAAGCGGCTAGGACAAAGAGAAAGTTTATCG 131  
Dy 252 AAGGCTGACGAGAGGAGGATGAGTTAAGCGGCTAGGACAAAGAGAAAGTTTATCG 193  
Qy 132 CGGCTGTGGCGGCTGCGACATACAAAGACCATGGGCGGCGGATCGGATCAGCAGAGGAA 191  
Dy 192 CGGCTGTGGCGGCTGCGACATACAAAGACCATGGGCGGCGGATCGGATCAGCAGAGGAA 133  
Qy 192 GGCCTGACGAGAGCGATGGGATGAGTACGCCATGGCTTCCTCCGCGGAGAGCTCCGCC 251  
Dy 132 GGCCTGACGAGAGCGATGGGATGAGTACGCCATGGCTTCCTCCGCGGAGAGCTCCGCC 73  
Qy 252 GGGGAGTACGAGTGGGAATATGACGAAGAGAGAGAGAAACCAGCTGGAGATTGAGAGA 311  
Dy 72 GGGGAGTACGAGTGGGAATATGACGAAGAGAGAGAGAAACCAGCTGGAGATTGAGAGA 13  
Qy 312 CTGGAGGAGCAG 323  
Dy 12 CTGGAGGAGCAG 1

## RESULT 15

ABA34138/C  
ID ABA34138 standard; DNA; 312 BP.

XX ABA34138;

XX 23-JAN-2002 (first entry)

XX Probe #12604 for gene expression analysis in human heart cell sample.

XX Human; gene expression; heart; microarray; vascular system; probe;  
XX cardiovascular disease; hypertension; cardiac arrhythmia;  
XX congenital heart disease; ss.

XX Homo sapiens.

XX WO200157274-A2.

XX 09-AUG-2001.

XX 30-JAN-2001; 2001WO-US00666.

XX 04-FEB-2000; 2000US-0180312.

XX 26-MAY-2000; 2000US-0207456.

XX 30-JUN-2000; 2000US-0608408.

XX 03-AUG-2000; 2000US-0632366.

XX 21-SEP-2000; 2000US-0234687.

XX 27-SEP-2000; 2000US-0236359.

XX 04-OCT-2000; 2000GB-0024263.

XX (MOLE-) MOLECULAR DYNAMICS INC.

XX Penn SG, Hanzel DK, Chen W, Rank DR;

XX WPI; 2001-48899/53.  
XX Single exon nucleic acid probes for analyzing gene expression in human  
PT hearts -  
XX  
XX  
PS Claim 4; SEQ ID No 12604; 530pp; English.  
XX  
XX The present invention relates to single exon nucleic acid probes for  
CC measuring human gene expression in a sample derived from human heart. The  
CC present sequence is one such probe. The probes may be used for  
CC predicting, measuring and displaying gene expression in samples derived  
CC from the human heart via microarrays. By measuring gene expression, the  
CC probes are useful for predicting, diagnosing, grading, staging,  
CC monitoring and prognosing diseases of the human heart and vascular system  
CC e.g. cardiovascular disease, hypertension, cardiac arrhythmias and  
CC congenital heart disease.  
CC Note: The sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences.  
XX  
SQ Sequence 312 BP; 41 A; 125 C; 66 G; 80 T; 0 other;

Query Match 8.2%; Score 310.4; DB 22; Length 312;  
Best Local Similarity 99.7%; Pred. No. 6.5e-74;  
Matches 311; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
QY 12 ATGGCGACTGCGCGCGAACCCTCGGCTTCAGAACCGAGGCTGAGTCCCAAGGCTGGGCC 71  
Db 312 ATGGCGACTGCGCGCGAACCCTCGGCTTCAGAACCGAGGCTGAGTCCCAAGGCTGGGCC 253  
QY 72 AAGGCTGACGAGGAGGAGGATGAGTTAAGGGGGTAGGACCAAGGAGAAAGGTGTATCG 131  
Db 252 AAGGCTGACGAGGAGGAGGATGAGTTAAGGGGGTAGGACCAAGGAGAAAGGTGTATCG 193  
QY 132 CGGGCTGTGGCGCTCGGCATACAGACCATGGGGCCAGCGTGGGATCAGCAGGAGAA 191  
Db 192 CGGGCTGTGGCGCTCGGCATACAGACCATGGGGCCAGCGTGGGATCAGCAGGAGAA 133  
QY 192 GCGCTGAGCGAGAGCGATGGGATGAGTACGCCATGGCTTCCTCCGCGAGAGCTCCCC 251  
Db 132 GCGCTGAGCGAGAGCGATGGGATGAGTACGCCATGGCTTCCTCCGCGAGAGCTCCCC 73  
QY 252 GGGGAGTACGAGTGGGAATATGACGAAGAGGAGGAGAGAAACACAGCTGGAGATTGAGAGA 311  
Db 72 GGGGAGTACGAGTGGGAATATGACGAAGAGGAGGAGAGAAACACAGCTGGAGATTGAGAGA 13  
QY 312 CTGGAGGAGCAG 323  
Db 12 CTGGAGGAGCAG 1

Search completed: June 22, 2003, 07:08:46  
Job time : 535 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 22, 2003, 06:58:01 ; Search time 127 seconds  
(without alignments)  
9171.326 Million cell updates/sec

Title: US-09-763-985A-1

Perfect score: 3798

Sequence: 1 ccacggtccgatggcgact.....tcaaaaaaaaaaaaaaaaaa 3798

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 441362 seqs, 15338381 residues

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Issued\_Patents\_NA:\*

1: /cgn2\_6/ptodata/1/ina/5A.COMB.seq:\*

2: /cgn2\_6/ptodata/1/ina/5B.COMB.seq:\*

3: /cgn2\_6/ptodata/1/ina/6A.COMB.seq:\*

4: /cgn2\_6/ptodata/1/ina/6B.COMB.seq:\*

5: /cgn2\_6/ptodata/1/ina/PCRTUS.COMB.seq:\*

6: /cgn2\_6/ptodata/1/ina/backfiles1.seq:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
C 1	95.6	2.5	7218	1	US-08-232-463-14
C 2	57.8	1.5	1926	4	US-09-249-585A-4
C 3	57.8	1.5	1931	2	US-09-130-114-2
C 4	51.4	1.4	436	4	US-09-397-787-257
C 5	49.4	1.3	51259	3	US-08-781-891-209
C 6	47.6	1.3	2277	1	US-08-676-967-2
C 7	47.6	1.3	2277	1	US-08-676-974-2
C 8	47.6	1.3	2277	2	US-09-098-487-2
C 9	47.4	1.2	289	4	US-09-007-005-17
C 10	47.4	1.2	289	4	US-09-244-796-17
C 11	47.4	1.2	1926	4	US-09-249-585A-2
C 12	47.4	1.2	2580	3	US-09-050-863-2
C 13	47.4	1.2	2580	4	US-09-359-081-2
C 14	47.4	1.2	5452	2	US-09-130-114-1
C 15	47.4	1.2	9600	4	US-08-910-647-1
C 16	47.4	1.2	9600	4	US-09-620-925-1
C 17	47.4	1.2	10596	1	US-07-884-811-15
C 18	47.4	1.2	10596	1	US-07-885-971-15
C 19	47.4	1.2	10596	1	US-08-087-783A-15
C 20	47.4	1.2	10596	1	US-08-194-088B-15
C 21	47.4	1.2	10596	2	US-08-194-087-15
C 22	47.4	1.2	10596	5	PCT-US93-04648-15
C 23	47.2	1.2	16442	3	US-08-781-891-208
C 24	47	1.2	2338	1	US-08-425-069-1
C 25	47	1.2	2338	2	US-08-317-844B-1
C 26	44.2	1.2	456	2	US-08-557-309B-16
C 27	44.2	1.2	456	3	US-08-834-306-16

28	44.2	1.2	456	4	US-08-993-674A-16
29	44.2	1.2	456	4	US-09-256-976-16
30	44.2	1.2	3489	2	US-08-728-323A-1
31	44.2	1.2	3489	4	US-09-298-568-1
C 32	44.2	1.2	32207	2	US-08-770-379-20
C 33	44.2	1.2	32207	4	US-08-757-669A-20
C 34	44.2	1.2	32207	4	US-09-230-371A-20
35	44	1.2	18443	4	US-09-078-294-6
36	42.4	1.1	282	4	US-09-461-697-205
37	42.4	1.1	306	4	US-09-461-697-203
38	42.4	1.1	696	4	US-09-461-697-193
39	42.4	1.1	699	4	US-09-461-697-191
40	42.4	1.1	717	4	US-09-461-697-189
41	42.4	1.1	774	4	US-09-461-697-187
42	42.4	1.1	819	4	US-09-461-697-185
43	42.4	1.1	1669	4	US-09-461-697-184
44	42.2	1.1	1052	1	US-08-466-603-1
45	42.2	1.1	1052	1	US-08-314-503A-1

#### ALIGNMENTS

RESULT 1  
US-08-232-463-14/c  
; Sequence 14, Application US/08232463  
; Patent No. 5670367  
; GENERAL INFORMATION:  
; APPLICANT: DORNER, F.  
; APPLICANT: SCHEIFLINGER, F.  
; APPLICANT: FALKNER, F. G.  
; TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS  
; NUMBER OF SEQUENCES: 52  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Foley & Lardner  
; STREET: 1800 Diagonal Road, Suite 500  
; CITY: Alexandria  
; STATE: VA  
; COUNTRY: USA  
; ZIP: 22313-0299  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/232,463  
; FILING DATE:  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US/07/935,313  
; FILING DATE:  
; APPLICATION NUMBER: EP 91 114 300.6  
; FILING DATE: 26-AUG-1991  
; ATTORNEY/AGENT INFORMATION:  
; NAME: BENT, Stephen A.  
; REGISTRATION NUMBER: 29,768  
; REFERENCE/DOCKET NUMBER: 30472/114 IMM  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (703)836-9300  
; TELEFAX: (703)683-4109  
; TELEX: 899149  
; INFORMATION FOR SEQ ID NO: 14:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 7218 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; IMMEDIATE SOURCE:  
; CLONE: pTZgpt-F1s  
US-08-232-463-14

Query Match

2.5%; Score 95.6; DB 1; Length 7218;













```
; FILING DATE: <Unknown>
; ATTORNEY/AGENT INFORMATION:
; NAME: Silva, Robin M.
; REGISTRATION NUMBER: 38,304
; REFERENCE/DOCKET NUMBER: A-65638/DJB/RMS
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 781-1989
; TELEFAX: (415) 949-8711
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2580 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: unknown
; TOPOLOGY: unknown
; MOLECULE TYPE: DNA
; SEQUENCE DESCRIPTION: SEQ ID NO: 2:
US-09-359-081-2

Query Match      1.2%; Score 47.4; DB 4; Length 2580;
Best Local Similarity 43.4%; Pred. No. 0.007;
Matches 219; Conservative 0; Mismatches 286; Indels 0; Gaps 0;

QY 42 GAACCCGAGGCTGAGTCCAAAGCTGGGCCCAAGCTGACGAGGAGGAGTGAAGTAAAG 101
Db 910 GAGCAGGAGGAGGGCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 969
QY 102 GCGGCTAGGCAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 161
Db 970 GAGCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1029
QY 162 ATGGGGCCAGGCTGGGATCAGCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 221
Db 1030 GAGGGCCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1089
QY 222 GCCATGGCTTCTCCCGGAGAGCTCCCGGGGAGTACGAGTGGGATATGACGAAGAG 281
Db 1090 GAGCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1149
QY 282 GAGGAGAAAACCACTGGAGATTGAGAGACTGGAGGAGGAGGAGGAGGAGGAGGAG 341
Db 1150 GAGGGCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1209
QY 342 GACTACAACTGCGATGTCATGATGATGATGATGATGATGATGATGATGATGATGAT 401
Db 1210 GAGGGCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1269
QY 402 GTGAGGATGGCCCGCAGAGATGATGATGATGATGATGATGATGATGATGATGATG 461
Db 1270 GAGCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1329
QY 462 GAGTGGCTGCATGACGAGATCAGCATGTCGCCAGGATGTCGCCAGGAGGAGGAGG 521
Db 1330 GGGCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1389
QY 522 GACCTCTTTGAGAAAGCCGTGAAGG 546
Db 1390 GCCGGGTCGAGGAGGTAGTGGAGG 1414

RESULT 14
US-09-130-114-1/c
; Sequence 1, Application US/09130114
; Patent No. 5976807
; GENERAL INFORMATION:
; APPLICANT: Horlick, Robert A.
; APPLICANT: Damaj, Bassam B.
; APPLICANT: Robbins, Alan K.
; TITLE OF INVENTION: Eukaryotic Cells Stably Expressing Genes
; FILE OF INVENTION: From Multiple Transfected Episomes
; FILE REFERENCE: 0867/ID903US1
; CURRENT APPLICATION NUMBER: US/09/130,114
; CURRENT FILING DATE: 1998-08-06
; NUMBER OF SEQ ID NOS: 36
```

```
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 1
; LENGTH: 5452
; TYPE: DNA
; ORGANISM: VEBNA
US-09-130-114-1

Query Match      1.2%; Score 47.4; DB 2; Length 5452;
Best Local Similarity 43.4%; Pred. No. 0.011;
Matches 219; Conservative 0; Mismatches 286; Indels 0; Gaps 0;

QY 42 GAACCCGAGGCTGAGTCCAAAGCTGGGCCCAAGCTGACGAGGAGGAGGAGTGAAGTAA 101
Db 1895 GAGCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1836
QY 102 GCGGCTAGGCAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 161
Db 1835 GAGCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1776
QY 162 ATGGGGCCAGGCTGGGATCAGCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 221
Db 1775 GAGGGCCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1716
QY 222 GCCATGGCTTCTCCCGGAGAGCTCCCGGGGAGTACGAGTGGGATATGACGAAGAG 281
Db 1715 GAGCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1656
QY 282 GAGGAGAAAACCACTGGAGATTGAGAGACTGGAGGAGGAGGAGGAGGAGGAGGAG 341
Db 1655 GAGGGCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1596
QY 342 GACTACAACTGCGATGTCATGATGATGATGATGATGATGATGATGATGATGATGAT 401
Db 1595 GAGGGCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1536
QY 402 GTGAGGATGGCCCGCAGAGATGATGATGATGATGATGATGATGATGATGATGATG 461
Db 1535 GAGCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1476
QY 462 GAGTGGCTGCATGACGAGATCAGCATGTCGCCAGGATGTCGCCAGGAGGAGGAG 521
Db 1475 GGGCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1416
QY 522 GACCTCTTTGAGAAAGCCGTGAAGG 546
Db 1415 GCCGGGTCGAGGAGGTAGTGGAGG 1391

RESULT 15
US-08-910-647-1
; Sequence 1, Application US/08910647
; Patent No. 6251433
; GENERAL INFORMATION:
; APPLICANT: Zuckermann et al.
; TITLE OF INVENTION: Compositions and Methods for
; TITLE OF INVENTION: Polynucleotide Delivery
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Chiron Corporation
; STREET: 4560 Horton Street
; CITY: Emeryville
; STATE: California
; COUNTRY: U.S.A.
; ZIP: 94608-2916
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/910,647
; FILING DATE:
; CLASSIFICATION: 514
```



GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 22, 2003, 08:55:51 ; Search time 346 Seconds  
(without alignments)

16107.770 Million cell updates/sec

Title: US-09-763-985A-1

Perfect score: 3798

Sequence: 1 ccacgctcgcagtgccgact.....tcaaaaaaaaaaaaaaaaaa 3798

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 1042519 seqs, 733713590 residues

Total number of hits satisfying chosen parameters: 2085038

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Published Applications NA:\*

- 1: /cgn2\_6/ptodata/2/pubpna/US07\_PUBCOMB.seq:\*
- 2: /cgn2\_6/ptodata/2/pubpna/PCT\_NEW\_PUB.seq:\*
- 3: /cgn2\_6/ptodata/2/pubpna/US06\_NEW\_PUB.seq:\*
- 4: /cgn2\_6/ptodata/2/pubpna/US06\_PUBCOMB.seq:\*
- 5: /cgn2\_6/ptodata/2/pubpna/US07\_NEW\_PUB.seq:\*
- 6: /cgn2\_6/ptodata/2/pubpna/PCTUS\_PUBCOMB.seq:\*
- 7: /cgn2\_6/ptodata/2/pubpna/US08\_NEW\_PUB.seq:\*
- 8: /cgn2\_6/ptodata/2/pubpna/US08\_PUBCOMB.seq:\*
- 9: /cgn2\_6/ptodata/2/pubpna/US09\_NEW\_PUB.seq:\*
- 10: /cgn2\_6/ptodata/2/pubpna/US09\_PUBCOMB.seq:\*
- 11: /cgn2\_6/ptodata/2/pubpna/US10\_NEW\_PUB.seq:\*
- 12: /cgn2\_6/ptodata/2/pubpna/US10\_PUBCOMB.seq:\*
- 13: /cgn2\_6/ptodata/2/pubpna/US60\_NEW\_PUB.seq:\*
- 14: /cgn2\_6/ptodata/2/pubpna/US60\_PUBCOMB.seq:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	2886	76.0	3537	10	US-09-821-883-30
2	2563.2	67.5	3355	10	US-09-925-300-556
3	1056.4	27.8	32145	9	US-09-764-891-7319
4	441.4	11.6	567	9	US-09-918-995-27590
5	393.4	10.4	432	10	US-09-867-701-1728
6	392.4	10.3	430	10	US-09-867-701-2028
7	310.4	8.2	312	10	US-09-864-761-19458
8	284	7.5	284	9	US-09-796-692-2624
9	284	7.5	284	9	US-10-040-862-2624
10	244	6.4	244	9	US-09-796-692-6818
11	244	6.4	244	9	US-10-040-862-6818
12	219.8	5.8	475	10	US-09-864-761-2742
13	129.6	3.4	185	10	US-09-864-761-18092
14	126.4	3.3	467	10	US-09-864-761-1331
15	55.6	1.5	272	10	US-09-864-761-22608
16	55.4	1.5	575	10	US-09-864-761-20733
17	55.4	1.5	1282	9	US-10-002-344A-89
18	55.4	1.5	1969	10	US-09-864-761-3972
19	54.8	1.4	554	9	US-10-101-487-69

20	54.8	1.4	554	9	US-10-101-487-106	Sequence 106, App
c 21	54.6	1.4	454	10	US-09-864-761-5848	Sequence 5848, App
22	51.4	1.4	436	10	US-09-876-889-257	Sequence 257, App
23	51.4	1.4	489	10	US-09-244-694-153	Sequence 153, App
24	51.4	1.4	494	9	US-09-918-995-29013	Sequence 29013, A
25	50.8	1.3	594	9	US-10-123-155-10	Sequence 10, Appl
26	50.4	1.3	522	9	US-10-101-487-71	Sequence 71, Appl
c 27	50.4	1.3	530	9	US-10-101-487-73	Sequence 73, Appl
c 28	49.8	1.3	659158	9	US-09-771-208-20	Sequence 20, Appl
c 29	49.8	1.3	1891139	9	US-10-067-514-1	Sequence 1, Appl
30	48.6	1.3	720	9	US-10-101-487-74	Sequence 74, Appl
c 31	48.6	1.3	720	9	US-10-101-487-76	Sequence 76, Appl
32	48	1.3	726	10	US-09-864-761-19707	Sequence 19707, A
33	48	1.3	1951	10	US-09-864-761-2926	Sequence 2926, Ap
34	47.8	1.3	299	10	US-09-864-761-21553	Sequence 21553, A
c 35	47.8	1.3	559	10	US-09-864-761-7684	Sequence 7684, Ap
c 36	47.8	1.3	8895	9	US-10-091-438-250	Sequence 250, App
c 37	47.8	1.3	8895	9	US-10-091-438-256	Sequence 256, App
c 38	47.8	1.3	8895	10	US-09-764-853-887	Sequence 887, App
c 39	47.8	1.3	8895	10	US-09-764-853-937	Sequence 937, App
c 40	47.8	1.3	9656	9	US-10-091-438-246	Sequence 246, App
c 41	47.8	1.3	9656	9	US-10-091-438-255	Sequence 255, App
c 42	47.8	1.3	9656	10	US-09-764-853-886	Sequence 886, App
c 43	47.8	1.3	9656	10	US-09-764-853-933	Sequence 933, App
c 44	47.4	1.2	8705	9	US-10-291-230-14	Sequence 14, Appl
45	47.2	1.2	484	10	US-09-864-761-2218	Sequence 2218, Ap

#### ALIGNMENTS

##### RESULT 1

US-09-821-883-30 Application US/09821883  
; Sequence 30, Application US/09821883  
; Patent NO. US20020061310A1

GENERAL INFORMATION:  
; APPLICANT: Laus, Reiner

; APPLICANT: Vidovic, Damir

; APPLICANT: Graddis, Thomas

; TITLE OF INVENTION: Compositions and Methods for Dendritic  
; TITLE OF INVENTION: Cell-Based Immunotherapy

; FILE REFERENCE: 7636-0022-30

; CURRENT APPLICATION NUMBER: US/09/821.883

; CURRENT FILING DATE: 2001-03-30

; PRIOR APPLICATION NUMBER: US 60/193,504

; PRIOR FILING DATE: 2000-03-30

; NUMBER OF SEQ ID NOS: 30

; SOFTWARE: FastSeq for Windows Version 4.0

; SEQ ID NO 30

; LENGTH: 3537

; TYPE: DNA

; ORGANISM: Artificial Sequence

; FEATURE:

; OTHER INFORMATION: SART-3-IC

US-09-821-883-30

Query Match	76.0%	Score 2886;	DB 10;	Length 3537;
Best Local Similarity	100.0%	Pred. No. 0;		
Matches 2886;	Conservative	0;	Mismatches	0;
			Indels	0;
			Gaps	0;
Oy	12	ATGCGCTGTCGGCGCGGAAACCTCGGCTTCAGAACCCGAGGCTGAGTCCAAAGGCTGGGCC 71		
Db	1	ATGCGGACTGCGGCGCGGAAACCTCGGCTTCAGAACCCGAGGCTGAGTCCAAAGGCTGGGCC 60		
Oy	72	AAGCTGACGAGAGGAGGATGAGTTAAGCGCGCTAGGACAAAGGAGAAAGTGTATCG 131		
Db	61	AAGCTGACGAGAGGAGGATGAGTTAAGCGCGCTAGGACAAAGGAGAAAGTGTATCG 120		
Oy	132	CGGCTGTGGCGCTGGGACATACAAAGACCATGGGCCAGCTGGGATCAGCAGGAGAA 191		
Db	121	CGGCTGTGGCGCTGGGACATACAAAGACCATGGGCCAGCTGGGATCAGCAGGAGAA 180		
Oy	192	GGCTGACGAGAGGAGGATGAGTACGCCATGGCTTCTCCCGGAGAGCTCCGCC 251		

Db	181	GGCGTGAGGAGCGCATGGGATGAGTACGCCATGGCTTCCTCCGCGGAGAGCTCCCCC	240	Db	1261	TGGCAGGCATACCTTGATTACCTGAGGAGAGGGTTGATTTTCAAAACAGACTCCAGTAAA	1320
QY	252	GGGGAGTACGAGTGGGAATATACGGAAGAGGAGGAGAAAACACAGCTGGAGATTGAGAGA	311	QY	1332	GAGCTGGAGAGCTTCAGGGCCGCCCTTACTCTGCTGCTTGGAGTATCTGAAGCAGAGGGTG	1391
Db	241	GGGGAGTACGAGTGGGAATATACGGAAGAGGAGGAGAAAACACAGCTGGAGATTGAGAGA	300	Db	1321	GAGCTGGAGAGCTTCAGGGCCGCCCTTACTCTGCTGCTTGGAGTATCTGAAGCAGAGGGTG	1380
QY	312	CTGGAGGACAGTGTCTATCAACGCTCTATGACTACAACTGCCATGTGGACTTGATCAGA	371	QY	1392	GAAGAGCGCTTCAATGAGAGTGGTGATCCAAGCTGCGTGATTATGACAGAACTGGGCTAGG	1451
Db	301	CTGGAGGACAGTGTCTATCAACGCTCTATGACTACAACTGCCATGTGGACTTGATCAGA	360	Db	1381	GAAGAGCGCTTCAATGAGAGTGGTGATCCAAGCTGCGTGATTATGACAGAACTGGGCTAGG	1440
QY	372	CTGCTCAGGCTGGAAGGGAGCTTACCAAGGTGAGGATGGCCGCCGAGAAAGATGAGTGAA	431	QY	1452	ATTGAGGCTCGACTGTGCAATAACTGACAGAAAGCTCGGGAACCTCGGGATGAGCATCATG	1511
Db	361	CTGCTCAGGCTGGAAGGGAGCTTACCAAGGTGAGGATGGCCGCCGAGAAAGATGAGTGAA	420	Db	1441	ATTGAGGCTCGACTGTGCAATAACTGACAGAAAGCTCGGGAACCTCGGGATGAGCATCATG	1500
QY	432	ATCTTTTCCCTTGACTGAAGAGCTCTGGCTGGAGTGGCTGCATGACGAGATCAGCATGGCC	491	QY	1512	ACCAGAGAAATGCCCAAGTACGCCAACATGTGGCTAGAGTATTACAACTCGGAAAGAGCT	1571
Db	421	ATCTTTTCCCTTGACTGAAGAGCTCTGGCTGGAGTGGCTGCATGACGAGATCAGCATGGCC	480	Db	1501	ACCAGAGAAATGCCCAAGTACGCCAACATGTGGCTAGAGTATTACAACTCGGAAAGAGCT	1560
QY	492	CAGGATGGCTGGACAGAGACAGCAGTGTATGACCTCTTTGAGAAAGCCGTGAAGATTAC	551	QY	1572	CATGCTGACACCCAGCAGCTGCCGGAAGGCTCTGCACCCGGGCGCTCAGTGCACCACTGAC	1631
Db	481	CAGGATGGCTGGACAGAGACAGCAGTGTATGACCTCTTTGAGAAAGCCGTGAAGATTAC	540	Db	1561	CATGCTGACACCCAGCAGCTGCCGGAAGGCTCTGCACCCGGGCGCTCAGTGCACCACTGAC	1620
QY	552	ATTTGCTCTAACATTTGGCTAGAGTATGCCAGTACTCATGTTGGTGGGATTTGGTCAGAAA	611	QY	1632	TACCCAGAGCAGCTCTGGAAGTGTACTCACCATGGAGAGGACAGAGAGTTCTTTAGAA	1691
Db	541	ATTTGCTCTAACATTTGGCTAGAGTATGCCAGTACTCATGTTGGTGGGATTTGGTCAGAAA	600	Db	1621	TACCCAGAGCAGCTCTGGAAGTGTACTCACCATGGAGAGGACAGAGAGTTCTTTAGAA	1680
QY	612	GGTGGCTTTGAGAAAGTTCCGCTCGCTGTTTGAAGGGCTCTCTGCTGCTGTTGGTTTACAT	671	QY	1692	GATTTGGGATATAGCTGTTTCAGAAAACCTGAAACCCGATTAGCTGCTCAATGAGCAGAGA	1751
Db	601	GGTGGCTTTGAGAAAGTTCCGCTCGCTGTTTGAAGGGCTCTCTGCTGCTGTTGGTTTACAT	660	Db	1681	GATTTGGGATATAGCTGTTTCAGAAAACCTGAAACCCGATTAGCTGCTCAATGAGCAGAGA	1740
QY	672	ATGACAAAGGACTCGCCCTCTGGGAGGCTTACCAGAGTGTGAAAGTCGAGTTGTGGAA	731	QY	1752	ATGAAAGGCTGCAGAGAAAGAAAGCAGCCCTTGTGCAGCAAGAAAGAAAGAAAGCTGAAACAA	1811
Db	661	ATGACAAAGGACTCGCCCTCTGGGAGGCTTACCAGAGTGTGAAAGTCGAGTTGTGGAA	720	Db	1741	ATGAAAGGCTGCAGAGAAAGAAAGCAGCCCTTGTGCAGCAAGAAAGAAAGAAAGCTGAAACAA	1800
QY	732	GCTGCTCGGCTTGAGAAAGTCCACAGTCTTTTCCGGGACAGTTGGGATCCCACTCAT	791	QY	1812	CGGAAAGAGCTCGGGCTGAGNAGAAAGCGTTTAAAGAAAGAAAGAAAGATCAGAGGCCCA	1871
Db	721	GCTGCTCGGCTTGAGAAAGTCCACAGTCTTTTCCGGGACAGTTGGGATCCCACTCAT	780	Db	1801	CGGAAAGAGCTCGGGCTGAGNAGAAAGCGTTTAAAGAAAGAAAGAAAGATCAGAGGCCCA	1860
QY	792	GATATGGAGGCCACATTTTCAGAGTATGAAGAAATGGTCAGAGACCCCAATACCAAGATCA	851	QY	1872	GAGAACCGCGAGCAGATGAGGACCATGAGAAAGAGTGGGGCGATGATGAAGAAAGAGAG	1931
Db	781	GATATGGAGGCCACATTTTCAGAGTATGAAGAAATGGTCAGAGACCCCAATACCAAGATCA	840	Db	1861	GAGAACCGCGAGCAGATGAGGACCATGAGAAAGAGTGGGGCGATGATGAAGAAAGAGAG	1920
QY	852	GTAATTCAGAACTATAACAAAGCCTACAGCAGCTGGAGAAATATAAACCCCTATGAAGAA	911	QY	1932	CTTCCAAACGAGAAAGGCTCGAGAACACATCCCTCGAGCTGGGAGAAACACAAATGTA	1991
Db	841	GTAATTCAGAACTATAACAAAGCCTACAGCAGCTGGAGAAATATAAACCCCTATGAAGAA	900	Db	1921	CTTCCAAACGAGAAAGGCTCGAGAACACATCCCTCGAGCTGGGAGAAACACAAATGTA	1980
QY	912	GCACCTTCCAGGCAGGACCAAGGCTGGCAGAAATATCAAGCATATATCGATTTTGAG	971	QY	1992	GAAGTAGCAGCAGGCGCCGCTGGGAAATGTCTGCCGTAGATGTGGAGCCCTTTCGAG	2051
Db	901	GCACCTTCCAGGCAGGACCAAGGCTGGCAGAAATATCAAGCATATATCGATTTTGAG	960	Db	1981	GAAGTAGCAGCAGGCGCCGCTGGGAAATGTCTGCCGTAGATGTGGAGCCCTTTCGAG	2040
QY	972	ATGAAAATTTGGCGATCTCTCGCATTCAGTTGATCTTTGAGCGCGCCCTGGTCGAGAAC	1031	QY	2052	CAGAAAGGAGAGGACGCTCCCTGAAGAGGGACATGCCCAAGGCTGCTGCACGACAGCAGC	2111
Db	961	ATGAAAATTTGGCGATCTCTCGCATTCAGTTGATCTTTGAGCGCGCCCTGGTCGAGAAC	1020	Db	2041	CAGAAAGGAGAGGACGCTCCCTGAAGAGGGACATGCCCAAGGCTGCTGCACGACAGCAGC	2100
QY	1032	TGCTTTGCCAGAACTATFGATTCGCTTACAGTACAGTACCTAGATCGACAACTGAAAGTA	1091	QY	2112	AAGGACAGCATCACCGTCTTTGTTCAGCAACCTGCCCCCTACAGCATCGAGAGCGCGACAG	2171
Db	1021	TGCTTTGCCAGAACTATFGATTCGCTTACAGTACAGTACCTAGATCGACAACTGAAAGTA	1080	Db	2101	AAGGACAGCATCACCGTCTTTGTTCAGCAACCTGCCCCCTACAGCATCGAGAGCGCGACAG	2160
QY	1092	AAGGATTTGGTTTTATCTGTACATAACCCGCTATTAGAACTGCCCTGGACAGTTGCC	1151	QY	2172	AGCTCAGGCCACTCTTCGAGGCCCTGTGGGAGGTTGGTCCAGATCCGACCATCTTCAGC	2231
Db	1081	AAGGATTTGGTTTTATCTGTACATAACCCGCTATTAGAACTGCCCTGGACAGTTGCC	1140	Db	2161	AGCTCAGGCCACTCTTCGAGGCCCTGTGGGAGGTTGGTCCAGATCCGACCATCTTCAGC	2220
QY	1152	TTATGAGTTCGCTACCTCTTGCCCATGGAGAGACATGGAGTTGATCATCAAGTAATTTCT	1211	QY	2232	AACCGTGGGGATTTCCGAGGTTACTGCTACGCTGGAGTTTAAAGAGAGAAATACGCCCTT	2291
Db	1141	TTATGAGTTCGCTACCTCTTGCCCATGGAGAGACATGGAGTTGATCATCAAGTAATTTCT	1200	Db	2221	AACCGTGGGGATTTCCGAGGTTACTGCTACGCTGGAGTTTAAAGAGAGAAATACGCCCTT	2280
QY	1212	GTAACCTTCGAGAAAGCTTTGAATGCCGCTTTCATCCAGGCCACTGATATGTGGAGATT	1271	QY	2292	CAGGCATCGAGATGGACCCGGAAAGTGTAGAAGGGAGGCCAATGTTGTTTCCCCCTGT	2351
Db	1201	GTAACCTTCGAGAAAGCTTTGAATGCCGCTTTCATCCAGGCCACTGATATGTGGAGATT	1260	Db	2281	CAGGCATCGAGATGGACCCGGAAAGTGTAGAAGGGAGGCCAATGTTGTTTCCCCCTGT	2340
QY	1272	TGCGAGGCATACCTTGATTACCTGAGGAGAGGGTTGATTTCAACAAGACTCCAGTAAA	1331	QY	2352	GTGGATAGACAAACCCCGATTTTAAAGTGTTCAGTACAGCATTCCCTAGAGAAA	2411
				Db	2341	GTGGATAGACAAACCCCGATTTTAAAGTGTTCAGTACAGCATTCCCTAGAGAAA	2400

QY	2412	CACAAGCTGTTCACTCAGGCCTGCCTTCTCTGTACTAAAGAGGAACCTAGAGAATC	2471
Db	2401	CACAAGCTGTTTCATCTCAGGCCTGCCTTCTCTGTACTAAAGAGGAACCTAGAGAATC	2460
QY	2472	TGTAAGGCTCATCGCACCGTGAAGGACCTCAGGCTGTGTACCAACCGGGCTGCCAACCA	2531
Db	2461	TGTAAGGCTCATCGCACCGTGAAGGACCTCAGGCTGTGTACCAACCGGGCTGCCAACCA	2520
QY	2532	AAGGGCTTGCCTACGTGGAGTATGAAATGAATCCAGGCGTCGCAAGGCTGTGATCAAG	2591
Db	2521	AAGGGCTTGCCTACGTGGAGTATGAAATGAATCCAGGCGTCGCAAGGCTGTGATCAAG	2580
QY	2592	ATGGACGGCATGACTATCAAGAGACATCATCAAGTGGCAATCAGCAACCTCTCTCAG	2651
Db	2581	ATGGACGGCATGACTATCAAGAGACATCATCAAGTGGCAATCAGCAACCTCTCTCAG	2640
QY	2652	AGCAAAAGTTCACAGAAAGCCAGAGACACAGGAAGGCACCAAGTGGCCCACTGCTTTTGCCG	2711
Db	2641	AGCAAAAGTTCACAGAAAGCCAGAGACACAGGAAGGCACCAAGTGGCCCACTGCTTTTGCCG	2700
QY	2712	CAGACATACGAGCGAGGGGGAAGGAAGACGACAGCTGTCTCTACTGCTCTGTCGCCCTG	2771
Db	2701	CAGACATACGAGCGAGGGGGAAGGAAGACGACAGCTGTCTCTACTGCTCTGTCGCCCTG	2760
QY	2772	CAGGCCCAAGTCTCTGCAGCTCTCTCAGGCTGACAAAGGCCCTCGCGGGGCTCTGCAGTT	2831
Db	2761	CAGGCCCAAGTCTCTGCAGCTCTCTCAGGCTGACAAAGGCCCTCGCGGGGCTCTGCAGTT	2820
QY	2832	GCGCCCCACGACGCCACCGAGGCACCAAGATGTCCAAATGCCGATTTTGCCACAGCTGTTT	2891
Db	2821	GCGCCCCACGACGCCACCGAGGCACCAAGATGTCCAAATGCCGATTTTGCCACAGCTGTTT	2880
QY	2892	CTGAGA	2897
Db	2881	CTGAGA	2886

## RESULT 2

```

US-09-925-300-556
; Sequence 556, Application US/09925300
; Patent No. US20020151681a1
; GENERAL INFORMATION:
; APPLICANT: Craig Rosen,
; APPLICANT: Steve Ruben
; TITLE OF INVENTION: Nucleic Acids, P
; FILE REFERENCE: PA101
; CURRENT APPLICATION NUMBER: US/09/925
; CURRENT FILING DATE: 2001-08-10
; PRIOR APPLICATION NUMBER: PCT/US00/0
; PRIOR FILING DATE: 2000-03-08
; PRIOR APPLICATION NUMBER: 60/124,270
; PRIOR FILING DATE: 1999-03-12
; NUMBER OF SEQ ID NOS: 1890
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 556

```

```

; LENGTH: 3355
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (210)
; OTHER INFORMATION: none
US-09-925-300-556

```

	Query Match	Best Local Similarity	67.5%;	Score 2563.2;	DB 10;	Length 3355;
	Matches 2576;	Conservative	99.8%;	Pred. No. 0;	Mismatches	0; Gaps 1;
Qy	1205	AATTCCTGTAACTTCGAGAAAGCTTTGAATGCGGGTTCATCCAGGCCACTGATTATGT				1264
Db	386	AATCCAGTAACCTTCGAGAAAGCTTTGAATGCGGGTTCATCCAGGCCACTGATTATGT				445

QY	1265	GGAGATTGGCAGGCATACCTTGATTACCTTGAGGAGAAGGGTTGATTTCAACAAGACTC	1324
DB	446	GGAGATTGGCAGGCATACCTTGATTACCTTGAGGAGAAGGGTTGATTTCAACAAGACTC	505
QY	1325	CAGTAAAGAGCTGGAGGAGTTGAGGGCCGCTTTACTCGTGCCTTGGAGTATCTGAAGCA	1384
DB	506	CAGTAAAGAGCTGGAGGAGTTGAGGGCCGCTTTACTCGTGCCTTGGAGTATCTGAAGCA	565
QY	1385	GGAGGTGGAAGAGCGTTTCAATGAGAGTGGTGATCCAAGCTGCGTGATTTAGCAGAACTG	1444
DB	566	GGAGGTGGAAGAGCGTTTCAATGAGAGTGGTGATCCAAGCTGCGTGATTTAGCAGAACTG	625
QY	1445	GGCTAGAGTTGAGGCTCGACTGTCGCAATAACATGCAGAAAGCTCGGGAACCTCGGGATAG	1504
DB	626	GGCTAGAGTTGAGGCTCGACTGTCGCAATAACATGCAGAAAGCTCGGGAACCTCGGGATAG	685
QY	1505	CATCATGACAGAGAGAAATGCCAAGTACGCCAACATGTGCTAGAGTATTACAACCTGGA	1564
DB	586	CATCATGACAGAGAGAAATGCCAAGTACGCCAACATGTGCTAGAGTATTACAACCTGGA	745
QY	1565	AAGAGCTCATGGTGACACCCAGCACCTGCCGGAAGGCTCTCACCGGCCCTCCAGTGCAC	1624
DB	746	AAGAGCTCATGGTGACACCCAGCACCTGCCGGAAGGCTCTCACCGGCCCTCCAGTGCAC	805
QY	1625	CAGTGACTACCCAGAGCAGCTCTCGGAAGTGTTTACTCACCATGGAGGAGACAGAAGTTTC	1684
DB	806	CAGTGACTACCCAGAGCAGCTCTCGGAAGTGTTTACTCACCATGGAGGAGACAGAAGTTTC	865
QY	1685	TTTAAAGATTGGGATATAGCTGTTTCAGAAAACCTGAAACCCGATTAGCTCGTGTCAATGA	1744
DB	866	TTTAAAGATTGGGATATAGCTGTTTCAGAAAACCTGAAACCCGATTAGCTCGTGTCAATGA	925
QY	1745	GCAGAGAATGAAGGCTCGAGAGAAGAGCAGCCCTTGTGCAGCAAGAAAGAAAGAGGC	1804
DB	926	GCAGAGAATGAAGGCTCGAGAGAAGAGCAGCCCTTGTGCAGCAAGAAAGAAAGAGGC	985
QY	1805	TGAACAACGGAAAAAGAGCTCGGGCTGAGAGAAAGCGTTTAAAAAGAAAGAAAGATCAG	1864
DB	986	TGAACAACGGAAAAAGAGCTCGGGCTGAGAGAAAGCGTTTAAAAAGAAAGAAAGATCAG	1045
QY	1865	AGGCCACAGAAAGCCGAGCAGATGAGGACGATGACAAAGTGGGGCCATGATGAAGA	1924
DB	1046	AGGCCACAGAAAGCCGAGCAGATGAGGACGATGACAAAGTGGGGCCATGATGAAGA	1105
QY	1925	AGAGCAGCTTTCAAAACGAGAGGGTTCGAGAACAGCATCCCTGCAGCTGGAGAAACACA	1984
DB	1106	AGAGCAGCTTTCAAAACGAGAGGGTTCGAGAACAGCATCCCTGCAGCTGGAGAAACACA	1165
QY	1985	AAATGTAGAAGTAGCAGAGGGCCGCTGGGAAATGTGTCGCGTAGATGTGAGGCCCCC	2044
DB	1166	AAATGTAGAAGTAGCAGAGGGCCGCTGGGAAATGTGTCGCGTAGATGTGAGGCCCCC	1225
QY	2045	TTCCAAGCAGAAGGAGGAGCGCTCCCTGAAGAGGGACATGCCCCAAGGTGCTGCACGA	2104
DB	1226	TTCCAAGCAGAAGGAGGAGCGCTCCCTGAAGAGGGACATGCCCCAAGGTGCTGCACGA	1285
QY	2105	CAGCAGCAAGGACAGCATCACCGTCTTTGTTCAGCAACCTGCGCTTACAGCATGCAGGAGCC	2164
DB	1286	CAGCAGCAAGGACAGCATCACCGTCTTTGTTCAGCAACCTGCGCTTACAGCATGCAGGAGCC	1345
QY	2165	GGACACGAAGCTCAGGCCACTCTTTCGAGGCGTGTGGGAGGTGGTCCAGATCCGACCCAT	2224
DB	1346	GGACACGAAGCTCAGGCCACTCTTTCGAGGCGTGTGGGAGGTGGTCCAGATCCGACCCAT	1405
QY	2225	CTTCAGCAACCGTGGGATTTCCAGGTTTACTGCTAGCTGGAGTTTAAAGAGAGAAATC	2284
DB	1406	CTTCAGCAACCGTGGGATTTCCAGGTTTACTGCTAGCTGGAGTTTAAAGAGAGAAATC	1465
QY	2285	AGCCCTTCAGGCACCTGGAGATGGACCGGAAAAGTGTAGAGGGAGGCCAATGTTTGTTC	2344
DB	1466	AGCCCTTCAGGCACCTGGAGATGGACCGGAAAAGTGTAGAGGGAGGCCAATGTTTGTTC	1525
QY	2345	CCCTGTGTGATTAAGACAAAACCCCGATTTTTAAAGGTGTTCAGGTACAGCACTTCCCT	2404







Db 121 AGCGAAACGAGAACACACCTGTTTTCGAGACACAGTTGTCCAAATGTTTCTGGCCAGCTC 180  
QY 3242 CGGCCCTTTTCTATGACACTTCTCTCCACCCTGCACAGCACATGTCGCCGTCATTCT 3301  
Db 181 CGGCCCTTTTCTATGACACTTCTCTCCACCCTGCACAGCACATGTCGCCGTCATTCT 240  
QY 3302 TTTAATTTAAAGATGAATGCGAGATGCTAGTAATTCACAGAATGGGCTCTGTGGGG 3361  
Db 241 TTTAATTTAAAGATGAATGCGAGATGCTAGTAATTCACAGAATGGGCTCTGTGGGG 300  
QY 3362 GTGGGTCTGAGGAATGACCTATATAAAACATTTGCTGGAGTTTGTTCATGAGGCGCTGTG 3421  
Db 301 GTGGGTCTGAGGAATGACCTATATAAAACATTTGCTGGAGTTTGTTCATGAGGCGCTGTG 360  
QY 3422 CATTGTTT-ATATTATGTTGTTGTAATGACATGTCAGCCCTGTTTTCATGTTTCCATAAA 3480  
Db 361 CATTGTTTAAATATGTTGTTGTAATGACATGTCAGCCCTGTTTTCATGTTTCCATAAA 420  
QY 3481 GCAGAATATTG 3492  
Db 421 GCAGAATATTG 432

RESULT 6

US-09-867-701-2028  
; Sequence 2028, Application US/09867701  
; Patent No. US20020132237A1  
; GENERAL INFORMATION:  
; APPLICANT: Aglate, Paul A.  
; APPLICANT: Jones, Robert  
; APPLICANT: Harlocker, Susan L.  
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY  
; FILE REFERENCE: 210121.497  
; CURRENT APPLICATION NUMBER: US/09/867,701  
; NUMBER OF SEQ ID NOS: 10912  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 2028  
; LENGTH: 430.  
; TYPE: DNA  
; ORGANISM: Homo sapien  
; FEATURE:  
; NAME/KEY: misc.feature  
; LOCATION: (1)...(430)  
; OTHER INFORMATION: n = A,T,C or G

Query Match 10.3%; Score 392.4; DB 10; Length 430;  
Best Local Similarity 98.8%; Pred. No. 3.5e-110;  
Matches 426; Conservative 0; Mismatches 2; Indels 3; Gaps 3;

QY 3064 AAGAGCCTTTCAGTCTCCTCATATGAGGCGAGTGGCAGAAAAGTGACCACTCTGCAG 3123  
Db 1 AAGAGCCTTTCAGTCTCCTCATATGAGGCGAGTGGCAGAAAAGTGACCACTCTGCAG 60  
QY 3124 GCTGGGCCAGGATGTTGTTCTGAGATAGTTTGTATCTTAAAGACTGAGGCACAGAA 3183  
Db 61 GCTGGG-CCAGATGTTGTTCTGAGATAGTTTGTATCTTAAAGACTGAGGCACAGAA 119  
QY 3184 GCGAAACGAGAACACA-CGTGTTTTTGTAGACAGTTGTCCAAATGTTTCTGCCAGCTCC 3242  
Db 120 GCGAAACGAGAACACACCTGTTTGTAGACACAGTTGTCCAAATGTTTCTGCCAGCTCC 179  
QY 3243 GCGCCCTTTTCTATGACACTTCTCTCCACCCTGCACAGCATGTGCCCTCATCTT 3302  
Db 180 GCGCCCTTTTGTATGACACTTCTCTCTCCACCCTGCACAGCATGTGCCCTCATCTT 239  
QY 3303 TTAATTTTAAAGATGAATGCGAGATGCTAGTAATTCACAGAATGGGCTCTGTGGGG 3362  
Db 240 TTAATTTTAAAGATGAATGCGAGATGCTAGTAATTCACAGAATGGGCTCTGTGGGG 299

QY 3363 TGGGTCTGAGGAAGTACGTATATAAACATTTTCTGGAGTTTGTTCATGAGGCGCTGTC 3422  
Db 300 TGGGTCTGAGGAAGTACGTATATAAACATTTTCTGGAGTTTGTTCATGAGGCGCTGTC 359  
QY 3423 ATTTT-ATATTATGTTTGTAAATGACATGTCAGCCCTGTTTTCATGTTTCCATAAAAG 3481  
Db 360 ATTTTAAATATGTTTGTAAATGACATGTCAGCCCTGTTTTCATGTTTCCATAAAAG 419  
QY 3482 CAGAATATTG 3492  
Db 420 CAGAATATTG 430

RESULT 7

US-09-864-761-19458/C  
; Sequence 19458, Application US/09864761  
; Patent No. US20020048763A1  
; GENERAL INFORMATION:  
; APPLICANT: Penn, Sharron G.  
; APPLICANT: Rank, David R.  
; APPLICANT: Hanzel, David K.  
; APPLICANT: Chen, Wensheng  
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR  
; FILE REFERENCE: Aeomica-X-1  
; CURRENT APPLICATION NUMBER: US/09/864,761  
; PRIOR FILING DATE: 2001-05-23  
; PRIOR APPLICATION NUMBER: US 60/180,312  
; PRIOR FILING DATE: 2000-02-04  
; PRIOR APPLICATION NUMBER: US 60/207,456  
; PRIOR FILING DATE: 2000-05-26  
; PRIOR APPLICATION NUMBER: US 09/632,366  
; PRIOR FILING DATE: 2000-08-03  
; PRIOR APPLICATION NUMBER: GB 24263.6  
; PRIOR FILING DATE: 2000-10-04  
; PRIOR APPLICATION NUMBER: US 60/236,359  
; PRIOR FILING DATE: 2000-09-27  
; PRIOR APPLICATION NUMBER: PCT/US01/00666  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00667  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00664  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00669  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00665  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00668  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00663  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00662  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00661  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00670  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: US 60/234,687  
; PRIOR FILING DATE: 2000-09-21  
; PRIOR APPLICATION NUMBER: US 09/608,408  
; PRIOR FILING DATE: 2000-06-30  
; PRIOR APPLICATION NUMBER: US 09/774,203  
; PRIOR FILING DATE: 2001-01-29  
; NUMBER OF SEQ ID NOS: 49117  
; SOFTWARE: Anomax Sequence Listing Engine vers. 1.1  
; SEQ ID NO 19458  
; LENGTH: 312  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; OTHER INFORMATION: MAP TO AC008119.6  
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 4.1  
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 4.9

OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 2.8  
OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 3.8  
OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 3.2  
OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 4.3  
OTHER INFORMATION: EXPRESSED IN B474, SIGNAL = 2.6  
OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 3.5  
OTHER INFORMATION: EXPRESSED IN HBL100, SIGNAL = 2.5  
OTHER INFORMATION: SWISSPROT HIT: P27476, EVALUE 1.10e+00  
OTHER INFORMATION: NT HIT: D63879.1, EVALUE 0.00e+00  
OTHER INFORMATION: EST\_HUMAN HIT: BF514110.1, EVALUE 0.00e+00  
US-09-864-761-19458

Query Match 8.2%; Score 310.4; DB 10; Length 312;  
Best Local Similarity 99.7%; Pred. No. 5.8e-85;  
Matches 311; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
QY 12 ATGGCGACTGCGGCGGAAACCTCGCTTCAGAACCCGAGGCTGAGTCCAAAGCTGGGCC 71  
Db 312 ATGGCGACTGCGGCGGAAACCTCGGCTTCAGAACCCGAGGCTGAGTCCAAAGCTGGGCC 253  
QY 72 AAGGCTGACGAGAGGAGGATGAGTTAAGCGGCTAGCAAGGAGAAAGGTGTATCG 131  
Db 252 AAGGCTGACGAGAGGAGGATGAGTTAAGCGGCTAGCAAGGAGAAAGGTGTATCG 193  
QY 132 CGGGCTGTGGCGCTGCGACATACAAAGACCATGGGCGGAGGATGGGATCAGCAGGAGAA 191  
Db 192 CGGGCTGTGGCGCTGCGACATACAAAGACCATGGGCGGAGGATGGGATCAGCAGGAGAA 133  
QY 192 GCGCTGACGAGAGGCGATGGGATGAGTACGCCATGGCTTCCTCGCGGAGAGCTCCGCC 251  
Db 132 GCGCTGACGAGAGGCGATGGGATGAGTACGCCATGGCTTCCTCGCGGAGAGCTCCGCC 73  
QY 252 GGGGAGTACGAGTGGGAATATGACCAAGAGGAGGAGAAACCAAGCTGAGATTGAGAGA 311  
Db 72 GGGGAGTACGAGTGGGAATATGACCAAGAGGAGGAGGAGAAACCAAGCTGAGATTGAGAGA 13  
QY 312 CTGGAGGAGCAG 323  
Db 12 CTGGAGGAGCAG 1

RESULT 8  
US-09-796-692-2624/c  
Sequence 2624, Application US/09796692  
Publication No. US20020198362A1  
GENERAL INFORMATION:  
APPLICANT: Gaiger, Alexander  
APPLICANT: Algate, Paul A.  
APPLICANT: Mannion, Jane  
APPLICANT: Retter, Marc  
APPLICANT: Corixa Corporation  
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE DETECTION, DIAGNOSIS AND THERAPY  
FILE REFERENCE: 2077 001200  
CURRENT APPLICATION NUMBER: US/09/796,692  
CURRENT FILING DATE: 2001-03-01  
PRIOR APPLICATION NUMBER: 60/186,126  
PRIOR FILING DATE: 2000-03-01  
PRIOR APPLICATION NUMBER: 60/190,479  
PRIOR FILING DATE: 2000-03-17  
PRIOR APPLICATION NUMBER: 60/200,545  
PRIOR FILING DATE: 2000-04-27  
PRIOR APPLICATION NUMBER: 60/200,303  
PRIOR FILING DATE: 2000-04-28  
PRIOR APPLICATION NUMBER: 60/200,779  
PRIOR FILING DATE: 2000-04-28  
PRIOR APPLICATION NUMBER: 60/200,999  
PRIOR FILING DATE: 2000-05-01  
PRIOR APPLICATION NUMBER: 60/202,084  
PRIOR FILING DATE: 2000-05-04  
PRIOR APPLICATION NUMBER: 60/206,201  
PRIOR FILING DATE: 2000-05-22  
PRIOR APPLICATION NUMBER: 60/218,950  
PRIOR FILING DATE: 2000-07-14

PRIOR APPLICATION NUMBER: 60/222,903  
PRIOR FILING DATE: 2000-08-03  
PRIOR APPLICATION NUMBER: 60/223,416  
PRIOR FILING DATE: 2000-08-04  
PRIOR APPLICATION NUMBER: 60/223,378  
PRIOR FILING DATE: 2000-08-07  
NUMBER OF SEQ ID NOS: 9597  
SOFTWARE: FastSeq for Windows Version 3.0  
SEQ ID NO 2624  
LENGTH: 284  
TYPE: DNA  
ORGANISM: Homo sapiens  
US-09-796-692-2624  
Query Match 7.5%; Score 284; DB 9; Length 284;  
Best Local Similarity 100.0%; Pred. No. 7.7e-77;  
Matches 284; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 297 CTGGAGATTGAGAGACTGGAGGAGCAGTTGTCTATCAACGCTCTATGACTTACCAAGTGAGGATGGCCGC 356  
Db 284 CTGGAGATTGAGAGACTGGAGGAGCAGTTGTCTATCAACGCTCTATGACTTACCAAGTGAGGATGGCCGC 225  
QY 357 GTGGACTTGCATCAGACTGCTCAGGCTGGAGGGGAGCTTACCAAGTGAGGATGGCCGC 416  
Db 224 GTGGACTTGCATCAGACTGCTCAGGCTGGAGGGGAGCTTACCAAGTGAGGATGGCCGC 165  
QY 417 CAGAAGATGAGTGAATCTTTCCCTTGTACTGAAGAGCTCTGGCTGGAGTGGCTGCAATGAC 476  
Db 164 CAGAAGATGAGTGAATCTTTCCCTTGTACTGAAGAGCTCTGGCTGGAGTGGCTGCAATGAC 105  
QY 477 GAGATCAGCATGCGCCAGGATGGCTTGGACAGAGACGCTGTATGACCTCTTTTGAGAAA 536  
Db 104 GAGATCAGCATGCGCCAGGATGGCTTGGACAGAGACGCTGTATGACCTCTTTTGAGAAA 45  
QY 537 GCGCTGAGGATTACATTGCTTACATTTGGCTAGAGATGAG 580  
Db 44 GCGCTGAGGATTACATTGCTTACATTTGGCTAGAGATGAG 1

RESULT 9  
US-10-040-862-2624/c  
Sequence 2624, Application US/10040862  
Publication No. US20030078396A1  
GENERAL INFORMATION:  
APPLICANT: Gaiger, Alexander  
APPLICANT: Algate, Paul A.  
APPLICANT: Mannion, Jane  
APPLICANT: Retter, Marc  
APPLICANT: Corixa Corporation  
TITLE OF INVENTION: Compositions and Methods for the Detection, Diagnosis and Ther  
FILE REFERENCE: 014058-013520US  
CURRENT APPLICATION NUMBER: US/10/040,862  
CURRENT FILING DATE: 2001-11-06  
PRIOR APPLICATION NUMBER: US 60/186,126  
PRIOR FILING DATE: 2000-03-01  
PRIOR APPLICATION NUMBER: US 60/190,479  
PRIOR FILING DATE: 2000-03-17  
PRIOR APPLICATION NUMBER: US 60/200,545  
PRIOR FILING DATE: 2000-04-27  
PRIOR APPLICATION NUMBER: US 60/200,303  
PRIOR FILING DATE: 2000-04-28  
PRIOR APPLICATION NUMBER: US 60/200,779  
PRIOR FILING DATE: 2000-04-28  
PRIOR APPLICATION NUMBER: US 60/200,999  
PRIOR FILING DATE: 2000-05-01  
PRIOR APPLICATION NUMBER: US 60/202,084  
PRIOR FILING DATE: 2000-05-04  
PRIOR APPLICATION NUMBER: US 60/206,201  
PRIOR FILING DATE: 2000-05-22  
PRIOR APPLICATION NUMBER: US 60/218,950  
PRIOR FILING DATE: 2000-07-14  
PRIOR APPLICATION NUMBER: US 60/222,903

;; PRIOR FILING DATE: 2000-08-03  
;; PRIOR APPLICATION NUMBER: US 60/223,416  
;; PRIOR FILING DATE: 2000-08-04  
;; PRIOR APPLICATION NUMBER: US 60/223,378  
;; PRIOR FILING DATE: 2000-08-07  
;; PRIOR APPLICATION NUMBER: US 09/796,692  
;; PRIOR FILING DATE: 2001-03-01  
;; NUMBER OF SEQ ID NOS: 10467  
;; SOFTWARE: FastSeq for Windows Version 3.0  
;; SEQ ID NO 2624  
;; LENGTH: 284  
;; TYPE: DNA  
;; ORGANISM: Homo sapiens  
US-10-040-862-2624

Query Match 7.5%; Score 284; DB 9; Length 284;  
Best Local Similarity 100.0%; Pred. No. 7.7e-77;  
Matches 284; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 297 CTGGAGATTGAGAGACTGGAGGAGCAGTTGTCTTATCAAGCTCTATGACTACAACTGCCAT 356  
Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
284 CTGGAGATTGAGAGACTGGAGGAGCAGTTGTCTTATCAAGCTCTATGACTACAACTGCCAT 225  
QY 357 GTGGACTTGTACAGACTGCTCAGGCTGGGAAGGGGAGCTTACCAAGTGGAGATGGCCCGC 416  
Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
224 GTGGACTTGTACAGACTGCTCAGGCTGGGAAGGGGAGCTTACCAAGTGGAGATGGCCCGC 165  
QY 417 CAGAAGATGAGTGAATCTTCCCTTGACTGAAGAGCTCTGGCTGGAGTGGCTGCATGAC 476  
Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
164 CAGAAGATGAGTGAATCTTCCCTTGACTGAAGAGCTCTGGCTGGAGTGGCTGCATGAC 105  
QY 477 GAGATCAGCATGCCCGCAGGATGCCCTGGACAGAGACAGCTGTATGACCTCTTTTGAGAAA 536  
Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
104 GAGATCAGCATGCCCGCAGGATGCCCTGGACAGAGACAGCTGTATGACCTCTTTTGAGAAA 45  
QY 537 GCCGTGAAGATTACATTTTCTCTACATTTGGCTAGATATGG 580  
Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
44 GCCGTGAAGATTACATTTTCTCTACATTTGGCTAGATATGG 1

RESULT 10  
US-09-796-692-6818/c  
; Sequence 6818, Application US/09796692  
; Publication No. US20020198362A1  
; GENERAL INFORMATION:  
; APPLICANT: Gaiger, Alexander  
; APPLICANT: Mannion, Jane  
; APPLICANT: Retter, Marc  
; APPLICANT: Corixa Corporation  
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE DETECTION, DIAGNOSIS AND THERAPY  
; FILE REFERENCE: 2077.001200  
; CURRENT FILING DATE: 2001-03-01  
; PRIOR APPLICATION NUMBER: US/09/796,692  
; PRIOR FILING DATE: 2000-03-01  
; PRIOR APPLICATION NUMBER: 60/186,126  
; PRIOR FILING DATE: 2000-03-01  
; PRIOR APPLICATION NUMBER: 60/190,479  
; PRIOR FILING DATE: 2000-03-17  
; PRIOR APPLICATION NUMBER: 60/200,545  
; PRIOR FILING DATE: 2000-04-27  
; PRIOR APPLICATION NUMBER: 60/200,303  
; PRIOR FILING DATE: 2000-04-28  
; PRIOR APPLICATION NUMBER: 60/200,779  
; PRIOR FILING DATE: 2000-04-28  
; PRIOR APPLICATION NUMBER: 60/200,999  
; PRIOR FILING DATE: 2000-05-01  
; PRIOR APPLICATION NUMBER: 60/202,084  
; PRIOR FILING DATE: 2000-05-04  
; PRIOR APPLICATION NUMBER: 60/206,201  
; PRIOR FILING DATE: 2000-05-22  
; PRIOR APPLICATION NUMBER: 60/218,950  
; PRIOR FILING DATE: 2000-07-14  
; PRIOR APPLICATION NUMBER: 60/222,903  
; PRIOR FILING DATE: 2000-08-03

;; PRIOR APPLICATION NUMBER: 60/223,416  
;; PRIOR FILING DATE: 2000-08-04  
;; PRIOR APPLICATION NUMBER: 60/223,378  
;; PRIOR FILING DATE: 2000-08-07  
;; NUMBER OF SEQ ID NOS: 9597  
;; SOFTWARE: FastSeq for Windows Version 3.0  
;; SEQ ID NO 6818  
;; LENGTH: 244  
;; TYPE: DNA  
;; ORGANISM: Homo sapiens  
US-09-796-692-6818  
Query Match 6.4%; Score 244; DB 9; Length 244;  
Best Local Similarity 100.0%; Pred. No. 1.5e-64;  
Matches 244; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 2991 CTTGGTGTCTGCTGCTCGCAACACAGATGGCTCCTCGGCTTTAGACAGAAAGGGAA 3050  
Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
244 CTTGGTGTCTGCTGCTCGCAACACAGATGGCTCCTCGGCTTTAGACAGAAAGGGAA 185  
QY 3051 GGGTTTCTAAGTCAAGAGCTTTTCACTGCTCCCTCATATTTAGGGCAGTGGCAGAAAAGT 3110  
Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
184 GGGTTTCTAAGTCAAGAGCTTTTCACTGCTCCCTCATATTTAGGGCAGTGGCAGAAAAGT 125  
QY 3111 GACCACTCTCAGGCTGGGCCAGGATGGTGTCTCTGAGATAGTTTTGTATCTTAAAGA 3170  
Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
124 GACCACTCTCAGGCTGGGCCAGGATGGTGTCTCTGAGATAGTTTTGTATCTTAAAGA 65  
QY 3171 CTGAGGCACAGAAAGCGAAGACAGACACACTGTTTTTGGACACACACTTGTCCAAATGTTT 3230  
Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
64 CTGAGGCACAGAAAGCGAAGACAGACACACTGTTTTTGGACACACACTTGTCCAAATGTTT 5  
QY 3231 CTGG 3234  
Db |||||  
4 CTGG 1  
RESULT 11  
US-10-040-862-6818/c  
; Sequence 6818, Application US/10040862  
; Publication No. US20030078396A1  
; GENERAL INFORMATION:  
; APPLICANT: Gaiger, Alexander  
; APPLICANT: Algate, Paul A.  
; APPLICANT: Mannion, Jane  
; APPLICANT: Retter, Marc  
; APPLICANT: Corixa Corporation  
; TITLE OF INVENTION: Compositions and Methods for the Detection, Diagnosis and Ther  
; FILE REFERENCE: 014058-013520US  
; CURRENT FILING DATE: 2001-11-06  
; PRIOR APPLICATION NUMBER: US/10/040,862  
; PRIOR FILING DATE: 2000-03-01  
; PRIOR APPLICATION NUMBER: US 60/186,126  
; PRIOR FILING DATE: 2000-03-17  
; PRIOR APPLICATION NUMBER: US 60/190,479  
; PRIOR FILING DATE: 2000-03-17  
; PRIOR APPLICATION NUMBER: US 60/200,545  
; PRIOR FILING DATE: 2000-04-27  
; PRIOR APPLICATION NUMBER: US 60/200,303  
; PRIOR FILING DATE: 2000-04-28  
; PRIOR APPLICATION NUMBER: US 60/200,779  
; PRIOR FILING DATE: 2000-04-28  
; PRIOR APPLICATION NUMBER: US 60/200,999  
; PRIOR FILING DATE: 2000-05-01  
; PRIOR APPLICATION NUMBER: US 60/202,084  
; PRIOR FILING DATE: 2000-05-04  
; PRIOR APPLICATION NUMBER: US 60/206,201  
; PRIOR FILING DATE: 2000-05-22  
; PRIOR APPLICATION NUMBER: US 60/218,950  
; PRIOR FILING DATE: 2000-07-14  
; PRIOR APPLICATION NUMBER: US 60/222,903  
; PRIOR FILING DATE: 2000-08-03  
; PRIOR APPLICATION NUMBER: US 60/223,416

;; PRIOR FILING DATE: 2000-08-04  
;; PRIOR APPLICATION NUMBER: US 60/223,378  
;; PRIOR FILING DATE: 2000-08-07  
;; PRIOR APPLICATION NUMBER: US 09/796,692  
;; PRIOR FILING DATE: 2001-03-01  
NUMBER OF SEQ ID NOS: 10467  
;; SOFTWARE: FastSeq for Windows Version 3.0  
;; SEQ ID NO 6818  
;; LENGTH: 244  
;; TYPE: DNA  
;; ORGANISM: Homo sapiens  
US-10-040-862-6818

Query Match 6.4%; Score 244; DB 9; Length 244;  
Best Local Similarity 100.0%; Pred. No. 1.5e-64;  
Matches 244; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2991 CCTGGTGTGCTGCTGCTGCAACACACAGATGGCTCTCGGCTTTAGACAGAAAGGGAA 3050  
Db 244 CCTGGTGTGCTGCTGCTGCAACACACAGATGGCTCTCGGCTTTAGACAGAAAGGGAA 185  
Qy 3051 GGGTTCCTAACTCAAGAGCCTTTTCAGTGCTCCCTCATATTCAGGCGAGTGGCAGAAAGT 3110  
Db 184 GGGTTCCTAACTCAAGAGCCTTTTCAGTGCTCCCTCATATTCAGGCGAGTGGCAGAAAGT 125  
Qy 3111 GACCACCTGTCAGGCTGGGCCAGGATGTGTGCTCCTGAGATAGTTTGTATCTTAAGA 3170  
Db 124 GACCACCTGTCAGGCTGGGCCAGGATGTGTGCTCCTGAGATAGTTTGTATCTTAAGA 65  
Qy 3171 CTGAGGCACAGAGCGAAACGAGACACACTGTTTTTGGACACACAGTTGTCCAAATGTTT 3230  
Db 64 CTGAGGCACAGAGCGAAACGAGACACACTGTTTTTGGACACACAGTTGTCCAAATGTTT 5  
Qy 3231 CTGG 3234  
Db 4 CTGG 1

RESULT 12

US-09-864-761-2742/c  
;; Sequence 2742, Application US/09864761  
;; Patent No. US20020048763A1  
;; GENERAL INFORMATION:  
;; APPLICANT: Penn, Sharron G.  
;; APPLICANT: Rank, David R.  
;; APPLICANT: Hanzel, David K.  
;; APPLICANT: Chen, Wensheng  
;; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR  
;; FILE REFERENCE: Aecomica-X-1  
;; CURRENT APPLICATION NUMBER: US/09/864,761  
;; CURRENT FILING DATE: 2001-05-23  
;; PRIOR APPLICATION NUMBER: US 60/180,312  
;; PRIOR FILING DATE: 2000-02-04  
;; PRIOR APPLICATION NUMBER: US 60/207,456  
;; PRIOR FILING DATE: 2000-05-26  
;; PRIOR APPLICATION NUMBER: US 09/632,366  
;; PRIOR FILING DATE: 2000-08-03  
;; PRIOR APPLICATION NUMBER: GB 24263.6  
;; PRIOR FILING DATE: 2000-10-04  
;; PRIOR APPLICATION NUMBER: US 60/236,359  
;; PRIOR FILING DATE: 2000-09-27  
;; PRIOR APPLICATION NUMBER: PCT/US01/00666  
;; PRIOR FILING DATE: 2001-01-30  
;; PRIOR APPLICATION NUMBER: PCT/US01/00667  
;; PRIOR FILING DATE: 2001-01-30  
;; PRIOR APPLICATION NUMBER: PCT/US01/00664  
;; PRIOR FILING DATE: 2001-01-30  
;; PRIOR APPLICATION NUMBER: PCT/US01/00669  
;; PRIOR FILING DATE: 2001-01-30  
;; PRIOR APPLICATION NUMBER: PCT/US01/00665  
;; PRIOR FILING DATE: 2001-01-30  
;; PRIOR APPLICATION NUMBER: PCT/US01/00668

;; PRIOR FILING DATE: 2001-01-30  
;; PRIOR APPLICATION NUMBER: PCT/US01/00663  
;; PRIOR FILING DATE: 2001-01-30  
;; PRIOR APPLICATION NUMBER: PCT/US01/00662  
;; PRIOR FILING DATE: 2001-01-30  
;; PRIOR APPLICATION NUMBER: PCT/US01/00661  
;; PRIOR FILING DATE: 2001-01-30  
;; PRIOR APPLICATION NUMBER: PCT/US01/00670  
;; PRIOR FILING DATE: 2001-01-30  
;; PRIOR APPLICATION NUMBER: US 60/234,687  
;; PRIOR FILING DATE: 2000-09-21  
;; PRIOR APPLICATION NUMBER: US 09/608,408  
;; PRIOR FILING DATE: 2000-06-30  
;; PRIOR APPLICATION NUMBER: US 09/774,203  
;; PRIOR FILING DATE: 2001-01-29  
NUMBER OF SEQ ID NOS: 49117  
;; SOFTWARE: Anomax Sequence Listing Engine vers. 1.1  
;; SEQ ID NO 2742  
;; LENGTH: 475  
;; TYPE: DNA  
;; ORGANISM: Homo sapiens  
;; FEATURE:  
;; OTHER INFORMATION: MAP TO AC008119.6  
;; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 4.1  
;; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 4.9  
;; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 2.8  
;; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 3.8  
;; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 3.2  
;; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 4.3  
;; OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 2.6  
;; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 3.5  
;; OTHER INFORMATION: EXPRESSED IN HBL100, SIGNAL = 2.5  
;; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 3.4  
US-09-864-761-2742

Query Match 5.8%; Score 219.8; DB 10; Length 475;  
Best Local Similarity 99.1%; Pred. No. 7.7e-57;  
Matches 221; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 104 GCCTAGGACAAGGAGAAAGGTCTTATCGCGGCTCTGCGGCTGCACATACAAGACCAT 163  
Db 475 GCCTAGGACAAGGAGAAAGGTCTTATCGCGGCTCTGCGGCTGCACATACAAGACCAT 416  
Qy 164 GGGCCAGCGTGGGATCAGCAGGAGGAGGCGCTGAGCGAGAGCGATGGGATGAGTACGC 223  
Db 415 GGGCCAGCGTGGGATCAGCAGGAGGAGGCGCTGAGCGAGAGCGATGGGATGAGTACGC 356  
Qy 224 CATGCTTCTCCCGGAGAGCTCCCGGGGAGTACGAGTGGGGAATATGACGAAGAGA 283  
Db 355 CATGCTTCTCCCGGAGAGCTCCCGGGGAGTACGAGTGGGGAATATGACGAAGAGA 296  
Qy 284 GGAGAAAACACAGCTGGAGATTGAGAGACTGGAGGAGCAGTTG 326  
Db 295 GGAGAAAACACAGCTGGAGATTGAGAGACTGGAGGAGCAGTTG 253

RESULT 13

US-09-864-761-18092/c  
;; Sequence 18092, Application US/09864761  
;; Patent No. US20020048763A1  
;; GENERAL INFORMATION:  
;; APPLICANT: Penn, Sharron G.  
;; APPLICANT: Rank, David R.  
;; APPLICANT: Hanzel, David K.  
;; APPLICANT: Chen, Wensheng  
;; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR  
;; FILE REFERENCE: Aecomica-X-1  
;; CURRENT APPLICATION NUMBER: US/09/864,761  
;; CURRENT FILING DATE: 2001-05-23  
;; PRIOR APPLICATION NUMBER: US 60/180,312  
;; PRIOR FILING DATE: 2000-02-04  
;; PRIOR APPLICATION NUMBER: US 60/207,456

```

: PRIOR FILING DATE: 2000-05-26
: PRIOR APPLICATION NUMBER: US 09/632,366
: PRIOR FILING DATE: 2000-08-03
: PRIOR APPLICATION NUMBER: GB 24263.6
: PRIOR FILING DATE: 2000-10-04
: PRIOR APPLICATION NUMBER: US 60/236,359
: PRIOR FILING DATE: 2000-09-27
: PRIOR APPLICATION NUMBER: PCT/US01/006666
: PRIOR FILING DATE: 2001-01-30
: PRIOR APPLICATION NUMBER: PCT/US01/006677
: PRIOR FILING DATE: 2001-01-30
: PRIOR APPLICATION NUMBER: PCT/US01/006684
: PRIOR FILING DATE: 2001-01-30
: PRIOR APPLICATION NUMBER: PCT/US01/006699
: PRIOR FILING DATE: 2001-01-30
: PRIOR APPLICATION NUMBER: PCT/US01/006655
: PRIOR FILING DATE: 2001-01-30
: PRIOR APPLICATION NUMBER: PCT/US01/006668
: PRIOR FILING DATE: 2001-01-30
: PRIOR APPLICATION NUMBER: PCT/US01/006663
: PRIOR FILING DATE: 2001-01-30
: PRIOR APPLICATION NUMBER: PCT/US01/006622
: PRIOR FILING DATE: 2000-09-21
: PRIOR APPLICATION NUMBER: US 09/608,408
: PRIOR FILING DATE: 2000-06-30
: PRIOR APPLICATION NUMBER: US 09/774,203
: PRIOR FILING DATE: 2001-01-29
: NUMBER OF SEQ ID NOS: 49117
: SOFTWARE: Annonax Sequence Listing Engine
: SEQ ID NO 18092
: LENGTH: 185
: TYPE: DNA
: ORGANISM: Homo sapiens
: FEATURE:
: OTHER INFORMATION: MAP TO AC008119.6
: OTHER INFORMATION: EXPRESSED IN ADULT L
: OTHER INFORMATION: EXPRESSED IN BRAIN,
: OTHER INFORMATION: EXPRESSED IN PLACENT,
: OTHER INFORMATION: EXPRESSED IN LUNG, S
: OTHER INFORMATION: EXPRESSED IN BONE MAR
: OTHER INFORMATION: EXPRESSED IN FETAL L
: OTHER INFORMATION: EXPRESSED IN BT474,
: OTHER INFORMATION: EXPRESSED IN HEART,
: OTHER INFORMATION: EXPRESSED IN HELA, S
: OTHER INFORMATION: EXPRESSED IN HEL100,
: OTHER INFORMATION: EST_HUMAN HIT: BEY39
: OTHER INFORMATION: SWISSPROT HIT: P0320
: OTHER INFORMATION: NT HIT: g17661951, E
: US-09-864-761-18092

```

[illegible]

```

Db
||||
4 ACTG 1

RESULT 14
US-09-864-761-1331/C
; Sequence 1331, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharron G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chan, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLE
; FILE REFERENCE: Aemica-X-1
; CURRENT APPLICATION NUMBER: US/09/864,761
; CURRENT FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: GB 24263.6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/006666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/006667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/006664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/006669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/006665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/006668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/006663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/006662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/006661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/006670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; PRIOR FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 49117
; SOFTWARE: Annonmax Sequence Listing Engine vers. 1.1
; SEQ ID NO 1331
; LENGTH: 467
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AC008119.6
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 1.6
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 1.2
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 1.9
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 1.4
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 1.2
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 1.8
; OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 1.5
; OTHER INFORMATION: EXPRESSED IN HEPA, SIGNAL = 1.6
; OTHER INFORMATION: EXPRESSED IN HBL100, SIGNAL = 1.9
; US-09-864-761-1331

```





**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 22, 2003, 06:49:56 ; Search time 3336 seconds  
(without alignments)  
18438.387 Million cell updates/sec

Title: US-09-763-985A-1  
Perfect score: 3798  
Sequence: 1 ccacgcgtccgatgcgcgct.....tcaaaaaaaaaaaaaaaaaa 3798

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 16154066 seqs, 8097743376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST:\*

1: em\_estba:\*  
2: em\_esthum:\*  
3: em\_estin:\*  
4: em\_estmu:\*  
5: em\_estov:\*  
6: em\_estpl:\*  
7: em\_estro:\*  
8: em\_htc:\*  
9: gb\_est1:\*  
10: gb\_est2:\*  
11: gb\_est3:\*  
12: gb\_est4:\*  
13: gb\_est5:\*  
14: gb\_est6:\*  
15: em\_estfun:\*  
16: em\_estom:\*  
17: gb\_gss:\*  
18: em\_gss\_hum:\*  
19: em\_gss\_inv:\*  
20: em\_gss\_pln:\*  
21: em\_gss\_vrt:\*  
22: em\_gss\_fun:\*  
23: em\_gss\_mam:\*  
24: em\_gss\_mus:\*  
25: em\_gss\_other:\*  
26: em\_gss\_pro:\*  
27: em\_gss\_rod:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1844.4	48.6	1865	11 BC032601	BC032601 Homo sapi
2	1752	46.1	1947	11 BC004547	BC004547 Homo sapi
3	1750.2	46.1	1938	11 BC024279	BC024279 Homo sapi
4	867.2	22.8	872	14 BQ231651	BQ231651 AGENCOURT
5	839.6	22.1	920	14 BQ430825	BQ430825 AGENCOURT
6	831.8	21.9	844	14 BQ223392	BQ223392 AGENCOURT

7	810.6	21.3	897	13 BM459745	BM459745 AGENCOURT
8	805.6	21.2	861	12 BG682796	BG682796 602651288
9	805.2	21.2	989	13 BM465186	BM465186 AGENCOURT
10	788.2	20.8	844	9 AL563618	AL563618 AL563618
11	779.8	20.5	843	9 AU121811	AU121811 AU121811
12	770	20.3	851	13 BI457408	BI457408 603185548
13	767.6	20.2	910	14 BQ242533	BQ242533 AGENCOURT
14	747.4	19.7	943	12 BG758037	BG758037 602715086
15	739.4	19.5	789	9 AU137933	AU137933 AU137933
16	738.4	19.4	801	12 BE739868	BE739868 601593154
17	738	19.4	855	14 BQ215978	BQ215978 AGENCOURT
18	731.8	19.3	761	9 AU125256	AU125256 AU125256
19	730.6	19.2	819	12 BG743313	BG743313 602634603
20	724.6	19.1	1209	12 BE902931	BE902931 601676649
21	722	19.0	737	14 BQ017983	BQ017983 UI-H-DP0-
22	721	19.0	793	12 BE733854	BE733854 601569239
23	714.2	18.8	759	9 AU131500	AU131500 AU131500
24	714	18.8	870	12 BE778264	BE778264 601463533
25	713.8	18.8	816	13 BI861041	BI861041 603391854
26	710.4	18.7	1030	12 BE731693	BE731693 601567092
27	709.6	18.7	725	10 AW173076	AW173076 x182b07.X
28	707.6	18.6	829	9 AL530012	AL530012 AL530012
29	705	18.6	870	9 AU124928	AU124928 AU124928
30	703.8	18.5	741	9 AU135222	AU135222 AU135222
31	700.2	18.4	708	14 BQ001342	BQ001342 UI-H-DH1-
32	697.4	18.4	850	12 BG254822	BG254822 602369252
33	695.4	18.3	807	9 AU143035	AU143035 AU143035
34	693.4	18.3	779	9 AU122296	AU122296 AU122296
35	693.2	18.3	862	12 BE794789	BE794789 601589239
36	693.2	18.3	928	12 BG284085	BG284085 602408307
37	672.2	17.7	756	10 BE410448	BE410448 601302031
38	672.2	17.7	823	13 BI222840	BI222840 602942581
39	664	17.5	664	14 BM972959	BM972959 UI-CF-EC1
40	658	17.3	712	12 BG683453	BG683453 602651288
41	657	17.3	692	12 BG682214	BG682214 602629521
42	652.6	17.2	931	12 BG253802	BG253802 602366687
43	642.8	16.9	650	14 BM771774	BM771774 K-EST0055
44	642	16.9	831	9 AU143744	AU143744 AU143744
45	641.4	16.9	655	12 BE796859	BE796859 601588170

#### ALIGNMENTS

RESULT 1  
BC032601  
LOCUS  
DEFINITION  
Homo sapiens, Similar to squamous cell carcinoma antigen recognised by T cells 3, clone IMAGE:5534381, mRNA.  
ACCESSION  
BC032601  
VERSION  
BC032601.1  
KEYWORDS  
HTC.  
SOURCE  
human.  
ORGANISM  
Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE  
1 (bases 1 to 1865)  
Strausberg, R.  
Direct Submission  
Submitted (06-JUN-2002) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA  
NIH-MGC Project URL: <http://mgc.nci.nih.gov>  
Contact: MGC help desk  
Email: [cgabbs-remail.nih.gov](mailto:cgabbs-remail.nih.gov)  
Tissue Procurement: ATCC  
CDNA Library Preparation: Life Technologies, Inc.  
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: National Institutes of Health Intramural Sequencing Center (NISC), Gaithersburg, Maryland  
Web site: <http://www.nisc.nih.gov/>  
REMARK  
COMMENT

Contact: nisc.mgc@nih.gov

Akhter, N., Ayelle, K., Beckstrom-Sternberg, S.M., Benjamin, B.,  
Blakesley, R.W., Bouffard, G.G., Breen, K., Brinkley, C., Brooks, S.,  
Dietrich, N.L., Granite, S., Guan, X., Gupta, J., Haghighi, P.,  
Hansen, N., Ho, S.-L., Karlins, E., Kwong, P., Laric, P., Legaspi, R.,  
Maduro, Q.L., Masiello, C., Maskeri, B., Mastrian, S.D., McCloskey, J.C.,  
McDowell, J., Pearson, R., Stantripop, S., Thomas, P.J., Touchman, J.W.,  
Tsurgeon, C., Vogt, J.L., Walker, M.A., Wetherby, K.D., Wiggins, L.,  
Young, A., Zhang, L.-H. and Green, E.D.

Clone distribution: MGC clone distribution information can be found  
through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>  
Series: IRAP Plate: 69 Row: k Column: 6  
This clone was selected for full length sequencing because it  
passed the following selection criteria: matched mRNA gi: 7661951  
This clone has the following problem: frame shifted.

## FEATURES

Location/Qualifiers

## source

1..1865

/organism="Homo sapiens"

/db\_xref="taxon:9606"

/clone="IMAGE:5534381"

/tissue="uterus, leiomyosarcoma"

/clone\_lib="NIH\_MGC\_71"

/lab\_host="DH10B"

/note="Vector: pCMV-SPORT6"

523 a 392 c 557 g 393 t

## BASE COUNT

## ORIGIN

Query Match 48.6%; Score 1844.4; DB 11; Length 1865;  
Best Local Similarity 99.7%; Pred. No. 0;  
Matches 1848; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 5 GCGTCGATGGCGACTGCGCGCGAAACCTCGCGCTTCAGAACCCGAGGCTGAGTCCAAAGGC 64  
DB 11  
DB 12 GCGCAAGATGGCGACTGCGCGCGAAACCTCGCGCTTCAGAACCCGAGGCTGAGTCCAAAGGC 71  
QY 65 TGGGCCCAAGGCTGACGAGAGAGAGATGAGTTAAGCGCGCTAGACAAAGAGAAAGT 124  
DB 72 TGGGCCCAAGGCTGACGAGAGAGAGATGAGTTAAGCGCGCTAGACAAAGAGAAAGT 131  
QY 125 GTTATCCGCGGCTGTGGCGGCTGCGACATACAAAGACCATGGCGCCAGCTGGGATCAGCA 184  
DB 132 GTTATCCGCGGCTGTGGCGGCTGCGACATACAAAGACCATGGCGCCAGCTGGGATCAGCA 191  
QY 185 GGAGGAAGCGGTGAGCGAGAGCGATGGGATGAGTACGCCATGGCTTCCTCCGCGGAGAG 244  
DB 192 GGAGGAAGCGGTGAGCGAGAGCGATGGGATGAGTACGCCATGGCTTCCTCCGCGGAGAG 251  
QY 245 CTCGCCCGGGAGTACGAGTGGGAATATGACGAAGAGGAGGAGAAACCCAGCTGGAGAT 304  
DB 252 CTCGCCCGGGAGTACGAGTGGGAATATGACGAAGAGGAGGAGAAACCCAGCTGGAGAT 311  
QY 305 TGAGAGACTGGAGGAGCAGTGTCTATCAACGCTATGACTACAACTGGCATGTGGACTT 364  
DB 312 TGAGAGACTGGAGGAGCAGTGTCTATCAACGCTATGACTACAACTGGCATGTGGACTT 371  
QY 365 GATCAGACTGCTCAGGCTGGAAGGGAGGCTTACCAAGGTGAGGATGGCCCGCCAGAAGAT 424  
DB 372 GATCAGACTGCTCAGGCTGGAAGGGAGGCTTACCAAGGTGAGGATGGCCCGCCAGAAGAT 431  
QY 425 GAGTGAATCTTCCCTTGACTGAAGAGCTGTGGCTGGAGTGGCTGCATGACGAGATCAG 484  
DB 432 GAGTGAATCTTCCCTTGACTGAAGAGCTGTGGCTGGAGTGGCTGCATGACGAGATCAG 491  
QY 485 CATGGCCCAAGGATGGCTGGACAGAGAGCAGCTGTATGACCTTTTGAGAAAGCGTGAA 544  
DB 492 CATGGCCCAAGGATGGCTGGACAGAGAGCAGCTGTATGACCTTTTGAGAAAGCGTGAA 551  
QY 545 GGATTACATTTGCTCCTAACATTTGGCTAGAGTATGGCCAGTACTCAGTTGGTGGGATGG 604  
DB 552 GGATTACATTTGCTCCTAACATTTGGCTAGAGTATGGCCAGTACTCAGTTGGTGGGATGG 611  
QY 605 TCAGAAAGGTGGCCCTTGAGAAAGTTCGCTCCGCTTTTGAAGGGCTCTCTGCTGTTGG 664

DB 612 TCAGAAAGGTGGCCCTTGAGAAAGTTCGCTCCGCTGTTGAAAGGGCTCTCTCGTCTGTGG 671  
QY 665 TTACATATGACCAAGGACTCGCCTCTGGAGGCTTACCAGAGAGTTTGAAGATGCGAT 724  
DB 672 TTACATATGACCAAGGACTCGCCTCTGGAGGCTTACCAGAGAGTTTGAAGATGCGAT 731  
QY 725 TGTGAAGCTGCTCGGCTTGAGAAAGTTCACACAGTCTTTTCCGCGGACAGTTCGGCATCC 784  
DB 732 TGTGAAGCTGCTCGGCTTGAGAAAGTTCACACAGTCTTTTCCGCGGACAGTTCGGCATCC 791  
QY 785 ACTCTATGATGAGGAGCCACATTTGACAGATGATGAAGATGATGAGAGAGCCCAATACC 844  
DB 792 ACTCTATGATGAGGAGCCACATTTGACAGATGATGAAGATGATGAGAGAGCCCAATACC 851  
QY 845 ACAGTCAGTAATTCAGAACTATAACAAAGCACTACAGCAGCTGGAGAAATATAACCCCTA 904  
DB 852 AGAGTCAGTAATTCAGAACTATAACAAAGCACTACAGCAGCTGGAGAAATATAACCCCTA 911  
QY 905 TGAAGAAGCACTGTTGACGAGAGAGCAACCAAGGCTGGCAGAAATATCAAGCATATATCGA 964  
DB 912 TGAAGAAGCACTGTTGACGAGAGAGCAACCAAGGCTGGCAGAAATATCAAGCATATATCGA 971  
QY 965 TTTTGAGATGAAATTTGGCGATCTGCTGCGATTCAGTTGATCTTTGAGCGGCCCTGGT 1024  
DB 972 TTTTGAGATGAAATTTGGCGATCTGCTGCGATTCAGTTGATCTTTGAGCGGCCCTGGT 1031  
QY 1025 CCAGAACTGCGCTTGTCCAGAGCTTATGGATCCGTTACAGTCACTACCTAGATCGACAAC 1084  
DB 1032 CCAGAACTGCGCTTGTCCAGAGCTTATGGATCCGTTACAGTCACTACCTAGATCGACAAC 1091  
QY 1085 GAAAGTAAAGGATTTGGTTTATCTGTACATACCGCGCTATTAGAAACTGCGCCCTGGAC 1144  
DB 1092 GAAAGTAAAGGATTTGGTTTATCTGTACATACCGCGCTATTAGAAACTGCGCCCTGGAC 1151  
QY 1145 AGTTGCTTATGGAGTCGGTACCTCTTGGCCATGAGAGACATGAGAGTTGATCATCAAGT 1204  
DB 1152 AGTTGCTTATGGAGTCGGTACCTCTTGGCCATGAGAGACATGAGAGTTGATCATCAAGT 1211  
QY 1205 AATTTCTGTAACTTCGAGAAAGCTTTGAATCCGCGCTTCATCCAGGCCACTGATTATGT 1264  
DB 1212 AATTTCTGTAACTTCGAGAAAGCTTTGAATCCGCGCTTCATCCAGGCCACTGATTATGT 1271  
QY 1265 GGAGATTTGGCAGGCAATACCTTTGATTAACCTGAGGAGAGGGTTGATTTCAACAAGACTC 1324  
DB 1272 GGAGATTTGGCAGGCAATACCTTTGATTAACCTGAGGAGAGGGTTGATTTCAACAAGACTC 1331  
QY 1325 CAGTAAAGAGCTGGAGGAGTTGAGGCGCCCTTTACTCGTCTGGAGTATCTGAAGCA 1384  
DB 1332 CAGTAAAGAGCTGGAGGAGTTGAGGCGCCCTTTACTCGTCTGGAGTATCTGAAGCA 1391  
QY 1385 GGAGTGAAGAGCGCTTTCAATGAGAGTGGTGCATCCAAAGCTCGTGCATTCAGAACTG 1444  
DB 1392 GGAGTGAAGAGCGCTTTCAATGAGAGTGGTGCATCCAAAGCTCGTGCATTCAGAACTG 1451  
QY 1445 GGCTAGGATTTAGGCTGAGCTGCAATAACATGAGAGAAAGCTCGGGAAGTCTGGGATAG 1504  
DB 1452 GGCTAGGATTTAGGCTGAGCTGCAATAACATGAGAGAAAGCTCGGGAAGTCTGGGATAG 1511  
QY 1505 CATCATGACGAGAGAAATGCCAAGTACGCCAAGCTGTGGCTAGAGTATTACAACCTGA 1564  
DB 1512 CATCATGACGAGAGAAATGCCAAGTACGCCAAGCTGTGGCTAGAGTATTACAACCTGA 1571  
QY 1565 AAGAGCTCATGTTGACACCCAGCACTGCGGGAAGCTCTGACCGGGCGCTCCAGTGCAC 1624  
DB 1572 AAGAGCTCATGTTGACACCCAGCACTGCGGGAAGCTCTGACCGGGCGCTCCAGTGCAC 1631  
QY 1625 CAGTCACTACCCAGAGCAGCTGCGAAAGTGTACTACCAATGAGAGAGACAGAGGTTTC 1684  
DB 1632 CAGTCACTACCCAGAGCAGCTGCGAAAGTGTACTACCAATGAGAGAGACAGAGGTTTC 1691  
QY 1685 TTTTAAAGATTTGGGATATAGCTGTTTCAGAAACTCAAAACCGATAGCTCGTGTCAATGA 1744



QY 1059 TACAGTCAGTACCTAGATGACAACTGAAAGTAAAGGATTTGGTTTATCTGTACATAAC 1118  
DB 1145 TACAGTCAGTACCTAGATGACAACTGAAAGTAAAGGATTTGGTTTATCTGTACATAAC 1204  
QY 1119 CGCGCTATTAGAACTGCCCTGGACAGTGGCTTATGAGTCGGTACCTCTGGCCATG 1178  
DB 1205 CGCGCTATTAGAACTGCCCTGGACAGTGGCTTATGAGTCGGTACCTCTGGCCATG 1264  
QY 1179 GAGACATGAGGCTGATCATCAAGTAATTTCTGTAACCTTCGAGAAAGCTTTGAATGCC 1238  
DB 1265 GAGACATGAGGCTGATCATCAAGTAATTTCTGTAACCTTCGAGAAAGCTTTGAATGCC 1324  
QY 1239 GGCTTCATCCAGCCACTGATTAATGAGGATTTGGAGCATACCTTGATACCTGAGG 1298  
DB 1325 GGCTTCATCCAGCCACTGATTAATGAGGATTTGGAGCATACCTTGATACCTGAGG 1384  
QY 1299 AGAAGGTTGATTTCAACAAGACTCCAGTAAAGACTGGAGAGTTGAGGCGGCTTT 1358  
DB 1385 AGAAGGTTGATTTCAACAAGACTCCAGTAAAGACTGGAGAGTTGAGGCGGCTTT 1444  
QY 1359 ACTCGTCCCTGGAGTATCTGAAGCAGAGGAGTGAAGAGCTTTCAATGAGAGTGGTGAT 1418  
DB 1445 ACTCGTCCCTGGAGTATCTGAAGCAGAGGAGTGAAGAGCTTTCAATGAGAGTGGTGAT 1504  
QY 1419 CCAAGCTCGGTGATTAATGAGAACTCGGTAGGATTTGAGCTCGACTGTGCAATAACATG 1478  
DB 1505 CCAAGCTCGGTGATTAATGAGAACTCGGTAGGATTTGAGCTCGACTGTGCAATAACATG 1564  
QY 1479 CAGAAAGCTCGGAAGCTCTGGATAGCATCATGACAGAGGAATGCCAAGTACGCCAAC 1538  
DB 1565 CAGAAAGCTCGGAAGCTCTGGATAGCATCATGACAGAGGAATGCCAAGTACGCCAAC 1624  
QY 1539 ATGTGCTAGAGTATTACAACCTGGAAGAGCTCATGTCACACCCAGCACTGCCGGAAG 1598  
DB 1625 ATGTGCTAGAGTATTACAACCTGGAAGAGCTCATGTCACACCCAGCACTGCCGGAAG 1684  
QY 1599 GCTCTCAGCGGGCGCTCAGTCGACCCAGTACTACCCAGAGCACGTCCTGCGAAGTGTTA 1658  
DB 1685 GCTCTCAGCGGGCGCTCAGTCGACCCAGTACTACCCAGAGCACGTCCTGCGAAGTGTTA 1744  
QY 1659 CTCACCATGGAGAGGACAGAAGTCTTTAGAGATTTGGGATATAGCTCTTCAGAAACT 1718  
DB 1745 CTCACCATGGAGAGGACAGAAGTCTTTAGAGATTTGGGATATAGCTCTTCAGAAACT 1804  
QY 1719 GAAACCCGATTAGCTCGTCTCAATGACAGAGAAATGAAGCTGCAGAGAAAGGAGCAGCC 1778  
DB 1805 GAAACCCGATTAGCTCGTCTCAATGACAGAGAAATGAAGCTGCAGAGAAAGGAGCAGCC 1864  
QY 1779 CTGTGTCAGAGAAAGAAAGGCTGAACACGGAAAGAGCTCGGCTCGAGAGAA 1838  
DB 1865 CTGTGTCAGAGAAAGAAAGGCTGAACACGGAAAGAGCTCGGCTCGAGAGAA 1924  
QY 1839 GCGTTAAAAAAGAAAAAGA 1860  
DB 1925 GCGTTAAAAAAGAAAAAGA 1946

RESULT 3  
LOCUS BC024279  
DEFINITION Homo sapiens, 1938 bp mRNA linear HTC 28-FEB-2002  
by T cells 3, clone IMAGE:4810833, mRNA.  
ACCESSION BC024279  
VERSION BC024279.1 GI:18999504  
KEYWORDS HTC.  
SOURCE Homo sapiens.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 1938)  
AUTHORS Strausberg, R.  
TITLE Direct Submission

## JOURNAL

REMARK  
COMMENT

Submitted (26-FEB-2002) National Institutes of Health, Mammalian  
Gene Collection (MGC), Cancer Genomics Office, National Cancer  
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,  
USA  
NIH-MGC Project URL: <http://mgc.nci.nih.gov>  
Contact: MGC help desk  
Email: [cgabbs-r@mail.nih.gov](mailto:cgabbs-r@mail.nih.gov)  
Tissue Procurement: ATCC  
cDNA Library Preparation: Ruben Laboratory  
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: Genome Sequence Centre,  
BC Cancer Agency, Vancouver, BC, Canada  
[info@bgsc.bc.ca](mailto:info@bgsc.bc.ca)  
Steven Jones, Jennifer Asano, Ian Bosdet, Yaron Butterfield,  
Susanna Chan, Readman Chiu, Chris Fjell, Erin Garland, Ran Guin,  
Leticia Hsiao, Martin Krzywinski, Reta Kutsche, Oliver Lee, Soo  
Sen Lee, Victor Ling, Carrie Mathewson, Candice McLeavy, Steven  
Ness, Pawan Pandoh, Anna-Liisa Prabhu, Parvaneh Saeedi, Jacqueline  
Schein, Duane Smalius, Michael Smith, Lorraine Spence, Jeff Stott,  
Michael Thorne, Miranada Tsai, Natasja van den Bosch, Jill Vardy,  
George Yang, Scott Zuyderduyn, Marco Marra.

Clone distribution: MGC clone distribution information can be found  
through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>  
Series: IRAL Plate: 42 Row: 9 Column: 8  
This clone was selected for full length sequencing because it  
passed the following selection criteria: matched mRNA gi: 7661951  
This clone has the following problem: frame shifted.

## FEATURES

## source

1. .1938  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:4810833"  
/tissue="IMAGE:4810833"  
/tissue\_type="Placenta, choriocarcinoma"  
/clone\_lib="NIH MGC 21"  
/lab\_host="DH10B-R"  
/note="Vector: pOTB7"  
BASE COUNT 544 a 412 c 567 g 415 t  
ORIGIN

Query Match 46.1%; Score 1750.2; DB 11; Length 1938;  
Best Local Similarity 95.4%; Pred. No. 0;  
Matches 1847; Conservative 0; Mismatches 8; Indels 82; Gaps 1;  
QY 6 CGTCCGATGGCGACTGCGGCCGAAACCTCGGCTTCAGAACCCGAGCTGAGTCCAAAGCT 65  
DB 1 CGCAAGATGGCGACTGCGGCCGAAACCTCGGCTTCAGAACCCGAGCTGAGTCCAAAGCT 60  
QY 66 GGGCCCAAGGCTGACGGAGAGGAGGATGAGTTAAGCGGCTAGGACAGGAGAAAGGTG 125  
DB 61 GGGCCCAAGGCTGACGGAGAGGAGGATGAGTTAAGCGGCTAGGACAGGAGAAAGGTG 120  
QY 126 TTATCGCGGGCTGTGGCGCTCGACATACAGACCATGGGCCAGCTGGGATCAGCAG 185  
DB 121 TTATCGCGGGCTGTGGCGCTCGACATACAGACCATGGGCCAGCTGGGATCAGCAG 180  
QY 186 GAGGAAGCGCTGAGCGAGAGCGATGGGATGAGTACGCCATGGCTTCTCCCGGAGAGC 245  
DB 181 GAGGAAGCGCTGAGCGAGAGCGATGGGATGAGTACGCCATGGCTTCTCCCGGAGAGC 240  
QY 246 TCCCGCGGGAGTACAGTGGGAATATGACGAGGAGGAGAGAGAGAGAGAGAGAT 305  
DB 241 TCCCGCGGGAGTACAGTGGGAATATGACGAGGAGGAGAGAGAGAGAGAGAT 300  
QY 306 GAGAGACTGGAGGAGCAGTTGTCTATCAACGCTATGACTACAACTGCCATGTGGACTTG 365  
DB 301 GAGAGACTGGAGGAGCAGTTGTCTATCAACGCTATGACTACAACTGCCATGTGGACTTG 360  
QY 366 ATCAGACTGCTCAGGCTGGAAGGGAGCTTACCAAGGTGAGGATGCCCGCCAGAGATG 425  
DB 361 ATCAGACTGCTCAGGCTGGAAGGGAGCTTACCAAGGTGAGGATGCCCGCCAGAGATG 420  
QY 426 AGTGAATCTTTCCCTTGACTGAAGAGCTCTGGCTGGAGTGGCTGCATGACGAGATCAGC 485









Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 897)

AUTHORS NIH-MGC http://mgi.nci.nih.gov/

TITLE National Institutes of Health, Mammalian Gene Collection (MGC)

JOURNAL Unpublished (1999)

COMMENT Contact: Robert Strausberg, Ph.D.

Email: cgapbs-r@mail.nih.gov

Tissue Procurement: ATCC

cDNA Library Preparation: Life Technologies, Inc.

cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)

DNA Sequencing by: Agencourt Bioscience Corporation

Clone Distribution: MGC clone distribution information can be

found through the I.M.A.G.E. Consortium/LLNL at:

http://image.llnl.gov

Plate: LLAM12220 row: m column: 06

High quality sequence stop: 744.

Location/Qualifiers

1. .897

/organism="Homo sapiens"

/db\_xref="taxon:9606"

/clone="IMAGE:5534381"

/clone\_lib="NIH\_MGC\_71"

/tissue\_type="leiomyosarcoma"

/lab\_host="DH10B (phage-resistant)"

/note="Organ: uterus; Vector: pCMV-SPORT6; Site\_1: NotI;

Site\_2: SalI; Cloned unidirectionally. Primer: Oligo dT.

Average insert size 2.1 kb.

235 a 191 c 289 g 182 t

BASE COUNT

ORIGIN

Query Match 21.3%; Score 810.6; DB 13; Length 897;

Best Local Similarity 96.3%; Pred. No. 3.4e-158;

Matches 852; Conservative 0; Mismatches 29; Indels 4; Gaps 2;

QY 6 CGTCCGATCGGACTCGCGCCCAACCTCGGCTTCAGAACCCGAGCTGATCCAGGCT 65

DB 1 CGCAAGATGGGACTCGCGCCCAACCTCGGCTTCAGAACCCGAGCTGATCCAGGCT 60

QY 66 GGGCCCAAGCTGACGAGGAGGATGAGTTAAGCGGCTAGGACAAAGGAGAAAGTG 125

DB 61 GGGCCCAAGCTGACGAGGAGGATGAGTTAAGCGGCTAGGACAAAGGAGAAAGTG 120

QY 126 TTATCGCGGCTGTGGCCGCTCGACATACAAAGACCATGGGCCAGCGTGGATCAGCAG 185

DB 121 TTATCGCGGCTGTGGCCGCTCGACATACAAAGACCATGGGCCAGCGTGGATCAGCAG 180

QY 186 GAGGAGGCGTGAGCGAGAGCGATGGGATGATGATGATGATGATGATGATGATGATG 245

DB 181 GAGGAGGCGTGAGCGAGAGCGATGGGATGATGATGATGATGATGATGATGATGATG 240

QY 246 TCCCGCGGAGTACGAGTGGGAATATGACGAGGAGGAGGAGGAGGAGGAGGAGGAGG 305

DB 241 TCCCGCGGAGTACGAGTGGGAATATGACGAGGAGGAGGAGGAGGAGGAGGAGGAGG 300

QY 306 GAGAGACTGGAGGAGCGTGTCTATCAACGCTCTATGACTACAACTGCCATGTGGACTTG 365

DB 301 GAGAGACTGGAGGAGCGTGTCTATCAACGCTCTATGACTACAACTGCCATGTGGACTTG 360

QY 366 ATCAGACTCTCAGCTGGAAGGAGCTTACCAGGTGAGGATGCCCGCCAGAGATG 425

DB 361 ATCAGACTCTCAGCTGGAAGGAGCTTACCAGGTGAGGATGCCCGCCAGAGATG 420

QY 426 ACTGAAATCTTCCCTTGACTCAAGAGCTCTCGCTGGAGTGGCTGCATGACGAGATCAGC 485

DB 421 ACTGAAATCTTCCCTTGACTCAAGAGCTCTCGCTGGAGTGGCTGCATGACGAGATCAGC 480

QY 486 ATGGCCAGGATGGCTGGACAGAGCAGCAGCTGTATGACCTCTTTGAGAAACCCGCTGAAG 545

DB 481 ATGGCCAGGATGGCTGGACAGAGCAGCAGCTGTATGACCTCTTTGAGAAACCCGCTGAAG 540

QY 546 GATTACATTTGCTTAACATTTGGCTAGAGTATGCCAGTACTACTGATGGTGGGATGGT 605

DB 541 GATTACATTTGCTTAACATTTGGCTAGAGTATGCCAGTACTACTGATGGTGGGATGGT 600

QY 606 CAGAAAGTGGCTTGAGAAAGTTCCGCTCCGCTGTTTGAAGGCTCTCTCGTCTGTGGT 665

DB 601 CAGAAAGTGGCTTGAGAAAGTTCCGCTCCGCTGTTTGAAGGCTCTCTCGTCTGTGGT 660

QY 666 TTACATATGACCAAAAGGACTCGCCCTCTCGGAGGCTTACCAGAGTTTCAAAGTGGGAT 725

DB 661 TTACATATGACCAAAAGGACTCGCCCTCTCGGAGGCTTACCAGAGTTTCAAAGTGGGAT 720

QY 726 GTGGAGTGGCTCGGCTTGAGAAAGTCCACAGTCTTTCCGGGAGAGTTGGGATCCCA 785

DB 721 GTGGAGTGGCTCGGCTTGAGAAAGTCCACAGTCTTTCCGGGAGAGTTGGGATCCCA 779

QY 786 CTCTATGATATGAGGAGGAGGACTTTCAGAGTATGAAGATGTCAGAGACCAATACCA 845

DB 780 CTCTATGATATGAGGAGGAGGACTTTCAGAGTATGAAGATGTCAGAGACCAATACCA 839

QY 846 GAG---TCAGTAATTCAGAACTATACAAAGGAGTACACAGCAGCTG 887

DB 840 CAGAGTCCAGTATTTTCAGAACTATACAAAGGAGTACACAGCAG 884

RESULT 8

LOCUS BG682796

DEFINITION BG682796 861 bp mRNA linear EST 01-MAY-2001

ACCESSION BG682796

VERSION BG682796.1 GI:13914193

KEYWORDS EST.

SOURCE human.

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 861)

NIH-MGC http://mgi.nci.nih.gov/

National Institutes of Health, Mammalian Gene Collection (MGC)

Unpublished (1999)

Contact: Robert Strausberg, Ph.D.

Email: cgapbs-r@mail.nih.gov

Tissue Procurement: ATCC

cDNA Library Preparation: Life Technologies, Inc.

cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)

DNA Sequencing by: Agencourt Bioscience Corporation

Clone Distribution: MGC clone distribution information can be

found through the I.M.A.G.E. Consortium/LLNL at:

http://image.llnl.gov

Plate: LLAM1615 row: p column: 22

High quality sequence stop: 856.

Location/Qualifiers

1. .861

/organism="Homo sapiens"

/db\_xref="taxon:9606"

/clone="IMAGE:4762245"

/clone\_lib="NIH\_MGC\_47"

/tissue\_type="neuroblastoma, cell line"

/lab\_host="DH10B (phage-resistant)"

/note="Organ: brain; Vector: pOTB7; Site\_1: XhoI; Site\_2:

EcoRI; cDNA made by oligo-dr priming. Directionally

cloned into EcoRI/XhoI sites using the following 5'

adaptor: GGCACGAG(G). Size-selected >500bp for average

insert size 1.8kb. Library constructed by Ling Hong in

the laboratory of Gerald M. Rubin (University of

California, Berkeley) using ZAP-cDNA synthesis kit

(Stratagene) and Superscript II RT (Life Technologies).

Note: this is a NIH\_MGC Library.

BASE COUNT 247 a 184 c 233 g 197 t

ORIGIN

Query Match 21.2%; Score 805.6; DB 12; Length 861;

Best Local Similarity 99.1%; Pred. No. 3.7e-157;

Matches 852; Conservative 0; Mismatches 4; Indels 4; Gaps 4;



```
||||| 584 ATTATGGAGATTGGCAGCATACCTGATTACCTGAGGAGGAGGTTGATTTCAAC 643
||||| 1318 AAGACTCCAGTAAGAGCTGGAGGAGTTGAGGCGCGCTTTACTCGTGCTTGGAGTATC 1377
||||| 644 AAGACTCCAGTAAGAGCTGGAGGAGTTGAGGCGCGCTTTACTCGTGCTTGGAGTATC 703
||||| 1378 TGAAGCAGGAGGTGGAAGAGCGTTTCAATGAGAGTGGTATCAAGCTCGGTGATTATGC 1437
||||| 704 TGAAGCAGGAGGTGGGAGAGCGTTTCAATGAGAGTGGTATCAAGCTCGGTGATTATGC 763
||||| 1438 AGAAGTGGGCTAGGATTGAGGCTCGACTGTGCAATACATGCAGAAAGCTCGGGAATCT 1497
||||| 764 AGAAGTGGGCTAGGATTGAGGCTCGACTGTGCAATACATGCAGAAAGCTCGGGAATCT 823
||||| 1498 GGGATAGCATCATGACCA--GAGGAAATGCCAAGTAC--GCCAATAGTGTGCTAGAGTATT 1554
||||| 824 GGGATAGCATCATGACCAAGAGAAATGCCAGTACCGCAACATGCTGGCTAGAGTATT 883
||||| 1555 ACAACCTGGAAA 1566
||||| 884 ACAACCTGGAA 895

RESULT 10
AL563618/c 844 bp mRNA linear EST 16-FEB-2001
LOCUS AL563618 LTI_NFL001_NBC4 Homo sapiens cDNA clone CS0DD009YB08 3
DEFINITION prime, mRNA sequence.
ACCESSION AL563618
VERSION AL563618.1 GI:12913189
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
TITLE 1 (bases 1 to 844)
JOURNAL Li, W.B., Gruber, C., Jessee, J. and Polayes, D.
COMMENT Full-length cDNA libraries and normalization
Unpublished (2001)
Contact: Genoscope
Genoscope - Centre National de Sequencage
BP 191 9106 Evry cedex - France
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr.
FEATURES
source
1..844
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="LTI_NFL001_NBC4"
/sex="male"
/tissue_type="neuroblastoma cells"
/lab_host="DH10B"
/note="Organ: brain; Vector: pCMVSPORT 6; 1st strand cDNA
was primed with a NotI-oligo(dT) primer. Five prime end
enriched, double-stranded cDNA was digested with Not I and
cloned into the Not I and Eco RV sites of the pCMVSPORT 6
vector. Library was normalized. Library was constructed
by Life Technologies. Contact : Feng Liang Life
Technologies, a division of Invitrogen 9800 Medical Center
Drive Rockville, Maryland 20850, USA Fax : (1) 301 610
8371 Email : fliang@lifetech.com URL :
http://fulllength.invitrogen.com"
BASE COUNT 216 a 215 c 202 g 191 t 20 others
ORIGIN

Query Match 20.8%; Score 788.2; DB 9; Length 844;
Best Local Similarity 95.7%; Pred. No. 1.5e-153;
Matches 808; Conservative 19; Mismatches 15; Indels 2; Gaps 2;
Y 2745 CAGCTCTCTACTGCTGCTGCTGCGCTGCGAGCCGAGGCTGCGAGCTCTCCTCAGGCTGAG 2804
||||| 844 CAGCTCTCTACTGCTGCTGCTGCGCTGCGAGCCGAGGCTGCGAGCTCTCCTCAGGCTGAG 785
```

Isogai, T.  
HRI human cDNA project  
Unpublished (2000)  
Contact: Takao Isogai  
Genomics Laboratory  
Helix Research Institute  
1532-3 Yana, Kisarazu, Chiba 292-0812, Japan  
Tel: 81-438-52-3975  
Fax: 81-438-52-3986  
Email: genomics@hri.co.jp  
HRI human cDNA project; 5'- & 3'-end one pass sequencing: Helix  
Research Institute; cDNA library construction: Department of  
Virology, Institute of Medical Science, University of Tokyo, and  
Helix Research Institute.

FEATURES  
Location/Qualifiers  
1..843  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="MAMMA1001023"  
/clone\_lib="MAMMA1"  
/tissue\_type="mammary gland"  
/note="Vector: pME18SF13"  
BASE COUNT 215 a 176 c 274 g 174 t 4 others  
ORIGIN

Query Match 20.5%; Score 779.8; DB 9; Length 843;  
Best Local Similarity 97.7%; Pred. No. 8.6e-152;  
Matches 820; Conservative 0; Mismatches 16; Indels 3; Gaps 3;

QY 3 ACGCGTCGATGGCGACTGCGCGCCGAAACCTCGCTTCAGAACCCGAGCTGAGTCCCAAG 62  
DB 6 AAGCGCAAGATGGCGACTGCGCGCGAAACCTCGCTTCAGAACCCGAGCTGAGTCCCAAG 65  
QY 63 GCTGGCCCAAGGCTGACGAGAGGAGGATGAGCTTAAGCGCGCTAGGACAAAGAGAAAG 122  
DB 66 GCTGGCCCAAGGCTGACGAGAGGAGGATGAGCTTAAGCGCGCTAGGACAAAGAGAAAG 125  
QY 123 GTGTATTCGCGGCTGTGGCGCTCGGACATACAGACCATGCGGCCAGCGCTGGGATCAG 182  
DB 126 GTGTATTCGCGGCTGTGGCGCTCGGACATACAGACCATGCGGCCAGCGCTGGGATCAG 185  
QY 183 CAGGAGGAAGCGGTGACGAGAGGAGGATGGGATGAGTACGCGCATGCGCTTCCTCCGCGGAG 242  
DB 186 CAGGAGGAAGCGGTGACGAGAGGAGGATGGGATGAGTACGCGCATGCGCTTCCTCCGCGGAG 245  
QY 243 AGCTCCCGCGGAGTACGAGTGGGAATATCACGAAGAGGAGGAGAAACCAAGCTGGAG 302  
DB 246 AGCTCCCGCGGAGTACGAGTGGGAATATCACGAAGAGGAGGAGAAACCAAGCTGGAG 305  
QY 303 ATTGAGAGCTGGAGGACAGTGTCTATCAACGCTCTATGACTACAACTGCCATGTGGAC 362  
DB 306 ATTGAGAGCTGGAGGACAGTGTCTATCAACGCTCTATGACTACAACTGCCATGTGGAC 365  
QY 363 TTGATCAGACTGCTCAGGCTGGAAGGAGGAGCTTACCAGAGTGGATGGCCCGCCAGAG 422  
DB 366 TTGATCAGACTGCTCAGGCTGGAAGGAGGAGCTTACCAGAGTGGATGGCCCGCCAGAG 425  
QY 423 ATGAGTGAATCTTTCCCTTGACTGAAGAGCTCTGGCTGGAGTGGCTGCATGACGAGATC 482  
DB 426 ATGAGTGAATCTTTCCCTTGACTGAAGAGCTCTGGCTGGAGTGGCTGCATGACGAGATC 485  
QY 483 AGCATGGCCCGAGGATGGCTGGACAGAGGACAGCTGTATGACCTCTTTTGAGAAAGCCGTG 542  
DB 486 AGCATGGCCCGAGGATGGCTGGACAGAGGACAGCTGTATGACCTCTTTTGAGAAAGCTGTG 545  
QY 543 AAGGATTACATTTGCTTAACATTTGGCTAGAGATGGCCAGTACTAGTTGGTGGGAT 602  
DB 546 AAGGATTACATTTGCTTAACATTTGGCTAGAGATGGCCAGTACTAATTTGGTGGGAT 605  
QY 603 GGTCAAGAGTGGCCTTGAGAAAGTTCGCTCCGTTGTTGAAGGGCTCTCTCGTCTCTT 662  
DB 606 GGTCAAGAGTGGCCTTGAGAAAGTTCGCTCCGTTGTTGAAGGGCTCTCTCGTCTCTT 665

QY 663 GGTTTACATATGACCAAGAGGACTCGCCCTCTGGAGGCTTACCAGAGAGTTTG-AAAGTGC 721  
DB 666 GGTTTACATATGACCAAGAGGACTCGCCCTCTGGAGGCTTACCAGAGAGTTTGAAAAGTGC 725  
QY 722 GATTGTGAAGCTGC-TCGGCTTGAGAAAGTCCACAGTCTTTTCCGGCGAGCTTGGCGA-780  
DB 726 CATTTGTGAAGCTGCTCGGCTTGAGAAAGTNCACAGTCTTTTCCGGCGAGCTTGGCGA-785  
QY 781 TCCACACTCTATGATATGAGGCGCCACATTTGCGAGAGTATGAAGATGTCAGAAAGACCCA 839  
DB 786 TCCC-TCTATGATATGAGGCGCCATTTGCGAGATATGAAGATGTCAGAAAGACCCA 843  
RESULT 12  
BI457408 851 bp mRNA linear EST 21-AUG-2001  
LOCUS 603185548F1 NIH\_MGC\_42 Homo sapiens cDNA clone IMAGE:5258362 5',  
DEFINITION mRNA sequence.  
ACCESSION BI457408  
VERSION BI457408.1 GI:15248064  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 851)  
AUTHORS NIH-MGC http://mgc.nci.nih.gov/  
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)  
JOURNAL Unpublished (1999)  
COMMENT Contact: Robert Strausberg, Ph.D.  
Email: cgapbs@mail.nih.gov  
Tissue Procurement: ATCC  
cDNA Library Preparation: Ling Hong/Rubin Laboratory  
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: Incyte Genomics, Inc.  
Clone Distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at:  
http://image.llnl.gov  
Plate: LCOM1869 row: p column: 11  
High quality sequence stop: 788.  
FEATURES  
Location/Qualifiers  
1..851  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:5258362"  
/clone\_lib="NIH\_MGC\_42"  
/tissue\_type="epithelioid carcinoma cell line"  
/lab\_host="DH10B (phage-resistant)"  
/note="Organ: pancreas; Vector: pOTB7; Site\_1: XhoI;  
Site\_2: EcoRI; cDNA made by oligo-dT priming.  
Directionally cloned into EcoRI/XhoI sites using the  
following 5' adaptor: GGCACGAG(G). Size-selected >500bp  
for average insert size 1.8kb. Library constructed by Ling  
Hong in the laboratory of Gerald M. Rubin (University of  
California, Berkeley) using ZAP-cDNA synthesis kit  
(Stratagene) and Superscript II RT (Life Technologies).  
Note: this is a NIH\_MGC Library. 1"  
BASE COUNT 226 a 166 c 284 g 175 t  
ORIGIN  
Query Match 20.3%; Score 770; DB 13; Length 851;  
Best Local Similarity 96.7%; Pred. No. 9.4e-150;  
Matches 818; Conservative 0; Mismatches 25; Indels 3; Gaps 3;  
QY 50 GGCTGAGTCCAAAGGCTGGCCCCCAAGGCTGACGGAGAGGAGGATGAGTTAAGCGCGTAG 109  
DB 1 GGCTGAGTCCAAAGGCTGGCCCCCAAGGCTGACGGAGAGGAGGATGAGTTAAGCGCGTAG 60  
QY 110 GACAAGGAGAAAGGTGTTATCGCGGCTGTGCGCCCTGCGACATACAAAGACCATGGGCC 169  
DB 61 GACAAGGAGAAAGGTGTTATCGCGGCTGTGCGCCCTGCGACATACAAAGACCATGGGCC 120  
QY 170 AGCGTGGGATCAGCAGGAGGAAGGCGCTGAGCGAGAGCGATGGGGATGAGTACGCCATGCC 229



```
Db 121 AGCGTGGGATCACAGAGAGAGCGCTGACGAGAGCGATGGGGATGAGTACCCCTGGC 180
QY 230 TTCTCTCCGGGAGAGCTCCCGGGGAGTACGAGTGGGAATATGACGAAGAGGAGAGAA 289
Db 181 TTCTCTCCGGGAGAGCTCCCGGGGAGTACGAGTGGGAATATGACGAAGAGGAGAGAA 240
QY 290 AAACCACTGGAGATTGAGAGACTGGAGGAGCAGTTGTCTATCAACGCTCTATGACTACAA 349
Db 241 AAAGCAGCTGGAGATTGAGAGACTGGAGGAGCAGTTGTCTATCAACGCTCTATGACTACAA 300
QY 350 CTGCCATGTGGACTTGATCAGACT-CCTCAGGCTGGAAGGGAGCTTACCAAGGTGAGGA 408
Db 301 CTGCCATGTGGACTTGATCAGACTGGCTCAGGCTGGAAGGGAGCTTACCAAGGTGAGGA 360
QY 409 TGGCCCGCCAGAGATGAGTGAATCTTTCCCTTGACTGAAGAGCTCTGGCTGGAGTGGC 468
Db 361 TGGCCCGCCAGAGATGAGTGAATCTTTCCCTTGACTGAAGAGCTCTGGCTGGAGTGGC 420
QY 469 TGCATGACGAGATCAGCATGGCCAGGATGGCTTGACAGAGCAGCAGTGTATGACTCT 528
Db 421 TGCATGACGAGATCAGCATGGCCAGGATGGCTTGACAGAGCAGCAGTGTATGACTCT 480
QY 529 TTGAGAAAGCCGTGAAGGATACATTTGCTTAAACATTTGGCTAGAGTATGGCCAGTACT 588
Db 481 TTGAGAAAGCCGTGAAGGATACATTTGCTTAAACATTTGGCTAGAGTATGGCCAGTACT 540
QY 589 CAGTTGGTGGGATTTGCTCAGAAA-GGTGGCTTTGAGAAAGTTCGCTCCGTTTGAAGG 647
Db 541 CAGTTGGTGGGATTTGCTCAGAAAAGGTGGCTTTGAGAAAGTTCGCTCCGTTTGAAGG 600
QY 648 GCTCTCTGCTGTGTGGTTTACATATGACCAAGGACTCGCCCTCTGGAGGCTTACCGA 707
Db 601 GCTCTCTGCTGTGTAGTTTAAACATATGACCAAGGACTCGCCCTCTGGAGGCTTACCGA 660
QY 708 GAG-TTTGAAAGTGCAATTTGGAAGCTGCTCGGCTTGAGAAAGTCCACAGTCTTTTCG 766
Db 661 GAGTTTGAAGTGCAATTTGGAAGCTGCTCGGCTTGAGAAAGTCCACAGTCTTTTCG 720
QY 767 GGCACAGTTTGGGATCCCACTCTATGATATGAGGCGCATTTGACAGATATGAAGAATG 826
Db 721 GCGAGAGTTTGGATACGATCTATGATATGAGGCGCAATTTGACAGATATGAAGAATG 780
QY 827 GTCAGAAAGCCAAATACAGAGTCACTAATCAGAACTATACAAAGCACTACAGAGCT 886
Db 781 GTCAGAAAGCCAAATACAGAGTCACTAATCAGAACTATACAAAGCACTTAAAGCAGTG 840
QY 887 GGAGAA 892
Db 841 TGAGAA 846
```

```
RESULT 13
BO924253
LOCUS BO924253
DEFINITION AGENCOURT_8865521 NIH_MGC_47 Homo sapiens cDNA clone IMAGE:6341844
5', mRNA sequence.
ACCESSION BO924253
VERSION BO924253.1 GI:22339284
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 910)
NIH-MGC http://mgi.nci.nih.gov/
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgabbs-remail.nih.gov
Tissue Procurement: ATCC
CDNA Library Preparation: Rubin Laboratory
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
```

```
FEATURES
source
1..910
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:6341844"
/lab_host="NIH_MGC_47"
/tissue_type="neuroblastoma, cell line"
/label="Organ: brain; Vector: pORF7; Site: 1: XhoI; Site: 2:
EcoRI; cDNA made by oligo-dT priming. Directionally
cloned into EcoRI/XhoI sites using the following 5'
adaptor: GGCACGAG(G). Size-selected >500bp for average
insert size 1.8kb. Library constructed by Ling Hong in
the laboratory of Gerald M. Rubin (University of
California, Berkeley) using ZAP-cDNA synthesis kit
(Stratagene) and Superscript II RT (Life Technologies).
Note: this is a NIH_MGC Library."
BASE COUNT 223 a 245 c 264 g 178 t
ORIGIN
Query Match 20.2%; Score 767.6; DB 14; Length 910;
Best Local Similarity 95.5%; Pred. No. 2.9e-149;
Matches 812; Conservative 0; Mismatches 34; Indels 4; Gaps 2;
2483 TGGCACCGGTGAAGGACCTCAGGCTGGTCCACCAACCGGGTGGCAACCAAGGCGCTGGC 2542
Db 1 TGGCACCGGTGAAGGACCTCAGGCTGGTCCACCAACCGGGTGGCAACCAAGGCGCTGGC 60
QY 2543 CTACGTGGAGTATGAATGAATCCAGCGCTCGAGGCTGTGATGAAGATGGACGCAT 2602
Db 61 CTACGTGGAGTATGAATGAATCCAGCGCTCGAGGCTGTGATGAAGATGGACGCAT 120
QY 2603 GACTATCAAGAGAGACATCATCAAGTGGCAATCAGCAACCTCTCTCAGAGAAAGTTCC 2662
Db 121 GACTATCAAGAGAGACATCATCAAGTGGCAATCAGCAACCTCTCTCAGAGAAAGTTCC 180
QY 2663 AGAGAGCCAGAGACCCAGGAGCCAGGTGGCCCCATGCTTTTGGCCGACACATACGG 2722
Db 181 AGAGAGCCAGAGACCCAGGAGCCAGGTGGCCCCATGCTTTTGGCCGACACATACGG 240
QY 2723 AGCGAGGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2782
Db 241 AGCGAGGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 300
QY 2783 TGCTGCAGCTCTCAGGCTGAGAGGCGCTCGCGGGCTCTCGAGTTGCCGCCCCAGC 2842
Db 301 TGCTGCAGCTCTCAGGCTGAGAGGCGCTCGCGGGCTCTCGAGTTGCCGCCCCAGC 360
QY 2843 AGCCACCGAGGACCCAGAGATCTCAATGCCGATTTTGGCAAGCTGTTTCTCAGAAAGTG 2902
Db 361 AGCCACCGAGGACCCAGAGATCTCAATGCCGATTTTGGCAAGCTGTTTCTCAGAAAGTG 420
QY 2903 AACGGGAGCTGGGAGACAGGAAATCCCTTACTTCTCTGCGCCGCGGAGCTCCACCC 2962
Db 421 AACGGGAGCTGGGAGACAGGAAATCCCTTACTTCTCTGCGCCGCGGAGCTCCACCC 480
QY 2963 ACCCAGCAGTGCACCTGGGAGTGGACAGGCGCTGGTGTGCTGTGCTGCGCAACACAGATG 3022
Db 481 ACCCAGCAGTGCACCTGGGAGTGGACAGGCGCTGGTGTGCTGTGCTGCGCAACACAGATG 540
QY 3023 GCTCTCGGCTTTAGACAGAAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 3082
Db 541 GCTCTCGGCTTTAGACAGAAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 600
QY 3083 CTCATATTGAGGCGAGTGGCAGAAAAGTGACCACTCTCGAGGCTGGGCCAGAGTGTGT 3142
Db 601 CTCATATTGAGGCGAGTGGCAGAAAAGTGACCACTCTCGAGGCTGGGCCAGAGTGTGT 660
```

DNA Sequencing by: Agencourt Bioscience Corporation  
Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at:  
<http://image.llnl.gov>  
Plate: LLC2540 row: i column: 13  
High quality sequence stop: 660.  
Location/Qualifiers







GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 18, 2003, 13:45:33 ; Search time 44 Seconds  
(without alignments)  
2916.372 Million cell updates/sec

Title: US-09-763-985A-2

Perfect score: 4994

Sequence: 1 MATAAETSAEPEAEKAGP.....AATEAPKMSNADFAKFLR 963

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A.Geneseq\_101002.\*

1: /SID52/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.\*  
2: /SID52/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.\*  
3: /SID52/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.\*  
4: /SID52/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.\*  
5: /SID52/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.\*  
6: /SID52/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.\*  
7: /SID52/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.\*  
8: /SID52/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.\*  
9: /SID52/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.\*  
10: /SID52/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.\*  
11: /SID52/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.\*  
12: /SID52/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.\*  
13: /SID52/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.\*  
14: /SID52/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.\*  
15: /SID52/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.\*  
16: /SID52/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.\*  
17: /SID52/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.\*  
18: /SID52/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.\*  
19: /SID52/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.\*  
20: /SID52/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.\*  
21: /SID52/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.\*  
22: /SID52/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.\*  
23: /SID52/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	4994	100.0	963	21	AA1985422
2	4989	99.9	1179	22	AA1913123
3	2918	58.4	578	21	AA1956918
4	526	10.5	104	22	AB1301113
5	526	10.5	104	22	AB135284
6	526	10.5	104	22	AB120728
7	526	10.5	104	22	AA156114
8	526	10.5	104	22	AA168486
9	526	10.5	104	22	AA16294
10	526	10.5	104	22	AA128783

11	526	10.5	104	22	AA104028
12	526	10.5	104	23	AB138067
13	513	10.3	108	23	AB133657
14	425	8.5	941	22	AB159801
15	420	8.4	100	21	AA101853
16	272	5.4	414	17	AA10529
17	272	5.4	414	22	AB130809
18	272	5.4	414	22	AB167385
19	268.5	5.4	702	22	AB159602
20	259.5	5.2	691	21	AA148233
21	259.5	5.2	705	21	AA148232
22	245.5	4.9	536	22	AA192836
23	245.5	4.9	707	16	AA179912
24	245.5	4.9	707	20	AA184052
25	245.5	4.9	707	22	AA148964
26	241.5	4.8	1009	22	AB158338
27	236.5	4.7	673	21	AA139359
28	236.5	4.7	683	21	AA139358
29	236	4.7	569	21	AA148234
30	234	4.7	836	22	AA140367
31	229.5	4.6	582	21	AA139360
32	229.5	4.6	650	21	AA152494
33	229.5	4.6	662	21	AA152493
34	229.5	4.6	675	21	AA152492
35	225	4.5	708	21	AA143567
36	224	4.5	483	14	AA143893
37	219.5	4.4	721	22	AB115463
38	217.5	4.4	1972	17	AA100024
39	217	4.3	41	22	AB128701
40	217	4.3	41	22	AB133883
41	217	4.3	41	22	AB119324
42	217	4.3	41	22	AA154649
43	217	4.3	41	22	AA167053
44	217	4.3	41	22	AA114915
45	217	4.3	41	22	AA127344

#### ALIGNMENTS

RESULT 1

AA1985422

ID AA1985422 standard; Protein; 963 AA.

AC AA1985422;

XX

DT 23-JUN-2000 (first entry)

XX

XX Tumour antigen protein SART-3.

DE Tumour antigen protein; SART-3; cytotoxic T-cell; HLA antigen; tumour.

XX Homo sapiens.

XX WO200012701-A1.

XX 09-MAR-2000.

XX 27-AUG-1999; 99WO-JP04622.

XX 28-AUG-1998; 98JP-0242660.

XX (SUMI) SUMITOMO PHARM CO LTD.

XX (ITOH) ITOH K.

XX Itoh K, Nakao M;

XX WPI; 2000-237868/20.

XX N-PSDB; AA290783.

XX Tumor antigen protein SART-3 recognized by cytotoxic T-cells binding to

XX HLA antigen for treatment and diagnosis of tumors

XX

PS	Claim 1; Page 54-58; 89pp; Japanese.
XX	This represents a tumour antigen protein (SART-3) which is recognized by cytotoxic T-cells binding to HLA antigen. SART-3 can be expressed by standard recombinant methodology. SART-3 and its peptide derivatives can be used in the prevention, treatment and diagnosis of tumours in vitro or in vivo.
XX	
SQ	Sequence 963 AA;
Query Match	100.0%; Score 4994; DB 21; Length 963;
Best Local Similarity	100.0%; Pred. No. 0;
Matches 963; Conservative	0; Mismatches 0; Indels 0; Gaps 0;
QY	1 MATAAETSASEPEAESKAGPKADGEDEVKAARTRRKVLRSRAVAATAATYKTMGPAMDQOEE 60
DB	1 MATAAETSASEPEAESKAGPKADGEDEVKAARTRRKVLRSRAVAATAATYKTMGPAMDQOEE 60
QY	61 GVSDEGDEYAMASSAESPGYEWYDEEEKNQLEIERLEEQLSINVDYNCHVDLIR 120
DB	61 GVSDEGDEYAMASSAESPGYEWYDEEEKNQLEIERLEEQLSINVDYNCHVDLIR 120
QY	121 LRLEGELTKVRMAROKMSEIPLTEELALEWLHDEISMAQDGLDREHYVDLFEKAVDY 180
DB	121 LRLEGELTKVRMAROKMSEIPLTEELALEWLHDEISMAQDGLDREHYVDLFEKAVDY 180
QY	181 ICPNIWLEXGOYSGVGIGOGKGLKVRVSFVERALSSVGLHMTKGLALWEAYREFESAIYE 240
DB	181 ICPNIWLEXGOYSGVGIGOGKGLKVRVSFVERALSSVGLHMTKGLALWEAYREFESAIYE 240
QY	241 AARLEKVSFLRRQLAIPLYDMEATPAEYEESEDDIPESVTQNYNKALQOQLEKYKPYEE 300
DB	241 AARLEKVSFLRRQLAIPLYDMEATPAEYEESEDDIPESVTQNYNKALQOQLEKYKPYEE 300
QY	301 ALLOAEAPRLAEYQVDFEMKIGDPARIQLIFERALVENCIVLPDLWIRYSQYLDRLQKV 360
DB	301 ALLOAEAPRLAEYQVDFEMKIGDPARIQLIFERALVENCIVLPDLWIRYSQYLDRLQKV 360
QY	361 KDLVLSVHNRAIRNCPTWALMSRYLLAMERHGVDRHOVISVTFEKNAGFTQATDYVEI 420
DB	361 KDLVLSVHNRAIRNCPTWALMSRYLLAMERHGVDRHOVISVTFEKNAGFTQATDYVEI 420
QY	421 WQAYLDLRRRVDFKODSKELEELRAAFTRALEYLKQVEERFNEGSDPSCVIMONWAR 480
DB	421 WQAYLDLRRRVDFKODSKELEELRAAFTRALEYLKQVEERFNEGSDPSCVIMONWAR 480
QY	481 IEARLCNNMOKARELWDSITWTRGNKAYANWLEYNLERAHGDTQCHRKALHRAVQCTSD 540
DB	481 IEARLCNNMOKARELWDSITWTRGNKAYANWLEYNLERAHGDTQCHRKALHRAVQCTSD 540
QY	541 YPEHVCEVLLTMEETEGSLEDWDIAVQKTETRLARVNEQRMKAKEAALVQOEEKAEQ 600
DB	541 YPEHVCEVLLTMEETEGSLEDWDIAVQKTETRLARVNEQRMKAKEAALVQOEEKAEQ 600
QY	601 RKRAREKALKKKKIRGPERKGADEDEKQWGDDEEQPSKRRRVENSIPAAGETQNY 660
DB	601 RKRAREKALKKKKIRGPERKGADEDEKQWGDDEEQPSKRRRVENSIPAAGETQNY 660
QY	661 EVAAGPAGCAADVVEPPSKQKEKASLRDMPKVLHDSKDSITVFNLSNLPYSMOEPT 720
DB	661 EVAAGPAGCAADVVEPPSKQKEKASLRDMPKVLHDSKDSITVFNLSNLPYSMOEPT 720
QY	721 KLRLFEACEGVVQIRPISNDRGDFRGYCYVEFEKESKALQALEMDRKSVEGRPMFVSPC 780
DB	721 KLRLFEACEGVVQIRPISNDRGDFRGYCYVEFEKESKALQALEMDRKSVEGRPMFVSPC 780
QY	781 VDKSNPDPKVPYRSTSLKHKLFISGLPFSCTKEELEICAHGTVKDLRLVTNRAGKP 840
DB	781 VDKSNPDPKVPYRSTSLKHKLFISGLPFSCTKEELEICAHGTVKDLRLVTNRAGKP 840
QY	841 KGLAYVEYENESQASQAVNMKGDMTKENIKVAINPDPORVPEKPTRKAPGPMLLP 900
DB	841 KGLAYVEYENESQASQAVNMKGDMTKENIKVAINPDPORVPEKPTRKAPGPMLLP 900

QY	901 QTYGARGKGTQLSLLPRALQRPSSAAAPQAENGPAAPAAVAPAAATEAPKMSNADFAKLF 960
DB	901 QTYGARGKGTQLSLLPRALQRPSSAAAPQAENGPAAPAAVAPAAATEAPKMSNADFAKLF 960
QY	961 LRK 963
DB	961 LRK 963

RESULT 2

AAE13123	ID	AAE13123	standard; Protein; 1179 AA.
XX	AC	AAE13123;	
XX	AC	AAE13123;	
DT	28-JAN-2002	(first entry)	
XX	Human SART3-IC-HER-2	membrane distal intracellular domain fusion protein.	
DE	Immunostimulatory fusion protein; IFP; antigen component; therapy;		
XX	immunostimulatory component; T-cell mediated immune response; DC;		
KW	dendritic cell; colon cancer; breast carcinoma; ovarian cancer;		
KW	human; HER-2 membrane distal intracellular domain; SART3-IC;		
KW	squamous cell carcinoma antigen; fusion protein.		
XX	Homo sapiens.		
OS	WO200174855-A2.		
XX	11-OCT-2001.		
PN	30-MAR-2001; 2001WO-US10515.		
XX	30-MAR-2000; 2000US-193504P.		
PR	(DEND-) DENDREON CORP.		
XX	Laus R, Vidovic D, Graddis T;		
PI	WPI: 2001-662965/76.		
XX	N-PSDB; AAD21574.		
DR	An immunostimulatory fusion protein comprising the intracellular domain		
XX	of HER-2 and an antigen elicits an immune response to the antigen and		
PT	is useful for the treatment of associated cancer associated .		
PS	Disclosure; Page 55-58; 59pp; English.		
XX	The invention relates to immunostimulatory fusion proteins (IFP) and		
CC	nucleic acid molecules encoding such proteins. The IFPs comprise a		
CC	polypeptide antigen component and an immunostimulatory component derived		
CC	from the intracellular domain of HER-2 protein which is effective to		
CC	elicit a protective dendritic cell (DC)-induced T-cell mediated cellular		
CC	immune response to the antigen. IFP or superactivated dendritic cells		
CC	are used to treat cancer e.g. breast carcinoma, ovarian and colon cancer		
CC	associated with a particularly antigen. The present sequence is a		
CC	fusion protein which comprises human squamous cell		
CC	carcinoma antigen, SART3-IC and mature human HER-2 membrane distal		
CC	intracellular domain.		
XX	Sequence 1179 AA;		
SQ	Query Match	99.9%; Score 4989; DB 22; Length 1179;	
	Best Local Similarity	100.0%; Pred. No. 0;	
	Matches 962; Conservative	0; Mismatches 0; Indels 0; Gaps 0;	
QY	1 MATAAETSASEPEAESKAGPKADGEDEVKAARTRRKVLRSRAVAATAATYKTMGPAMDQOEE 60		
DB	1 MATAAETSASEPEAESKAGPKADGEDEVKAARTRRKVLRSRAVAATAATYKTMGPAMDQOEE 60		
QY	61 GVSDEGDEYAMASSAESPGYEWYDEEEKNQLEIERLEEQLSINVDYNCHVDLIR 120		

Db 61 GVSDEGDEYAMASSAESPGYEWYDEEKEKQLEIERLEEQLSINVYDYNCHVDLIR 120  
QY 121 LRLREGELTKVMAKQKSEIFPLTEELWLEHLDEISMAODGLDREHVYDLFEKAVXDY 180  
Db 121 LRLREGELTKVMAKQKSEIFPLTEELWLEHLDEISMAODGLDREHVYDLFEKAVXDY 180  
QY 181 ICPNLTWEYGOYVSGGIGOKGLEKVRSVFERALSSVGLHMTKGLALWEAYREFESAIVE 240  
Db 181 ICPNLTWEYGOYVSGGIGOKGLEKVRSVFERALSSVGLHMTKGLALWEAYREFESAIVE 240  
QY 241 AARLEKVSHPERQLAIPLYDMEATFAEYEEHSEDPPIESVIONYNKALQLEKYPYEE 300  
Db 241 AARLEKVSHPERQLAIPLYDMEATFAEYEEHSEDPPIESVIONYNKALQLEKYPYEE 300  
QY 301 ALLOAEAPRLAEYQAYIDFEMKIGDPARIQILFERALVENCILVDPDLWIRYSOYLDROQLKV 360  
Db 301 ALLOAEAPRLAEYQAYIDFEMKIGDPARIQILFERALVENCILVDPDLWIRYSOYLDROQLKV 360  
QY 361 KDVLVSVHNRAIRNCPWTVALWSRYLLAMERHGVHDQVTSVTFKALNAGFTQATDYVEI 420  
Db 361 KDVLVSVHNRAIRNCPWTVALWSRYLLAMERHGVHDQVTSVTFKALNAGFTQATDYVEI 420  
QY 421 WQAYLDYLRRRVDFKODSSKELEELRAAFTRALEYLKOEVERFNESGDPSCVIMQNWAR 480  
Db 421 WQAYLDYLRRRVDFKODSSKELEELRAAFTRALEYLKOEVERFNESGDPSCVIMQNWAR 480  
QY 481 IEARLCNNMOKARELWDSIMTRGNKAYANMWLEYNLERAHGDTQHCRAKALHRAVQCTSD 540  
Db 481 IEARLCNNMOKARELWDSIMTRGNKAYANMWLEYNLERAHGDTQHCRAKALHRAVQCTSD 540  
QY 541 YPEHVEVLLTWTERTEGSLEDWDIAVQKTETRLARVNEQRMKAAKEAALVQOEEKEAQ 600  
Db 541 YPEHVEVLLTWTERTEGSLEDWDIAVQKTETRLARVNEQRMKAAKEAALVQOEEKEAQ 600  
QY 601 RKRAREKALKKKKIRGPEKRGADDEDEKEWGDDEEOPSKRRRVENSIPAAGETQNV 660  
Db 601 RKRAREKALKKKKIRGPEKRGADDEDEKEWGDDEEOPSKRRRVENSIPAAGETQNV 660  
QY 661 EVAAGPAGCAAVDVEPPSKQKAAKSLKRDMPKVLHDSSKDSITVFFVSNLPSYMOEPT 720  
Db 661 EVAAGPAGCAAVDVEPPSKQKAAKSLKRDMPKVLHDSSKDSITVFFVSNLPSYMOEPT 720  
QY 721 KLRLPEAGGEVWQIRPISNRGDFRGYCYVEFEKESALQALEMDRKSVEGRPMFVSPC 780  
Db 721 KLRLPEAGGEVWQIRPISNRGDFRGYCYVEFEKESALQALEMDRKSVEGRPMFVSPC 780  
QY 781 VDKSNPDKVFRYSTSLEKHLFTISGLPFSCTKEELEICKAHGTVKDLRLVTNRAGKP 840  
Db 781 VDKSNPDKVFRYSTSLEKHLFTISGLPFSCTKEELEICKAHGTVKDLRLVTNRAGKP 840  
QY 841 KGLAYVEYENESQAQVAMKMDGMTIKENIKVAISNPPORVPEKPTRKAPGGMILLP 900  
Db 841 KGLAYVEYENESQAQVAMKMDGMTIKENIKVAISNPPORVPEKPTRKAPGGMILLP 900  
QY 901 QTYGARGKRTQLSLLPRALQPSAAAPQAEANGPAAAPAAVAPAAATEAPKMSNADFAKLF 960  
Db 901 QTYGARGKRTQLSLLPRALQPSAAAPQAEANGPAAAPAAVAPAAATEAPKMSNADFAKLF 960  
QY 961 LR 962  
Db 961 LR 962

RESULT 3

AAB56918 standard; Protein; 578 AA.

XX AC

XX AAB56918;

XX 13-MAR-2001 (first entry)

XX Human prostate cancer antigen protein sequence SEQ ID NO:1496.

XX

KW Human; prostate cancer; prostate cancer antigen; detection; diagnosis;  
KW neuroprotective; cytostatic; cardioactive; immunomodulatory; muscular;  
KW vulnary; gastrointestinal; nephrotropic; antiinfective; gynaecological;  
KW antibacterial; gene therapy; neural; immune; reproductive; renal;  
KW gastrointestinal; pulmonary; cardiovascular; proliferative disorder;  
KW wound; infectious disease.  
OS Homo sapiens.  
XX  
PN WO200055174-A1.  
XX  
PD 21-SEP-2000.  
XX  
PF 08-MAR-2000; 2000WO-US05988.  
XX  
PR 12-MAR-1999; 99US-0124270.  
XX  
PA (HUMA-) HUMAN GENOME SCI INC.  
PA (ROSE/) ROSEN C A.  
XX  
PI Rosen CA, Ruben SM;  
XX  
DR WPI; 2000-587513/55.  
DR N-PSDB; AAF16121.  
XX  
PT Prostate cancer associated gene sequences, referred to as prostate  
PT cancer antigens, useful for treatment, prevention, and diagnosis of  
PT disorders such as prostate cancer -  
XX  
PS Claim 11; Page 1934-1936; 2338pp; English.  
XX  
CC AAF15566 to AAF16505 encode the human prostate cancer associated  
CC proteins, called prostate cancer antigens, given in AAB56363 to AAB57302.  
CC The prostate cancer antigens can have neuroprotective, cytostatic,  
CC cardioactive, immunomodulatory, muscular, vulnary, gastrointestinal,  
CC nephrotropic, antiinfective, gynaecological and antibacterial activities,  
CC and can be used in gene therapy. The prostate cancer antigen  
CC polynucleotides may be used for detection of prostate cancer, chromosome  
CC identification, as chromosome markers, and for numerous other diagnostic  
CC or research purposes. The prostate cancer antigens may be used to treat  
CC disorders such as neural, immune, muscular, reproductive,  
CC gastrointestinal, pulmonary, cardiovascular, renal, and proliferative  
CC disorders, wounds, and infectious diseases. AAF16506 to AAF16514 to  
CC AAB57303 represent sequences used in the exemplification of the present  
CC invention.  
XX  
SQ Sequence 578 AA;  
Query Match 58.4%; Score 2918; DB 21; Length 578;  
Best Local Similarity 99.1%; Pred. No. 5e-223;  
Matches 564; Conservative 0; Mismatches 5; Indels 0; Gaps 0;  
QY 395 DHQVTSVTFEKAALNAGFIQATDYVEIWAQYLDYLRVRVDFKODSSKELEELRAAFTRALE 454  
Db 10 DFPSTPVTPEKAALNAGFIQATDYVEIWAQYLDYLRVRVDFKODSSKELEELRAAFTRALE 69  
QY 455 YLKQVEEERFNEGSDPSCVIMONWARIARLNNMOKARELWDSIMTRNNAKYANMWLEY 514  
Db 70 YLKQVEEERFNEGSDPSCVIMONWARIARLNNMOKARELWDSIMTRNNAKYANMWLEY 129  
QY 515 YNLERAHGDTQHCRAKALHRAVQCTSDYPEHVCEVLLTMBERTEGSLEDWDIAVQKTETRLA 574  
Db 130 YNLERAHGDTQHCRAKALHRAVQCTSDYPEHVCEVLLTMBERTEGSLEDWDIAVQKTETRLA 189  
QY 575 RVNEQRMKAAKEAALVQOEEKEAQKRAAEKALKKKKIRGPEKRGADDEDEKEW 634  
Db 190 RVNEQRMKAAKEAALVQOEEKEAQKRAAEKALKKKKIRGPEKRGADDEDEKEW 249  
QY 635 DDEEOPSKRRRVENSIPAAGETQNVVEAAGPAGCAAVDVEPPSKQKAAKSLKRDMPK 694  
Db 250 DDEEOPSKRRRVENSIPAAGETQNVVEAAGPAGCAAVDVEPPSKQKAAKSLKRDMPK 309  
QY 695 VLHDSSKDSITVFFVSNLPSYMOEPTDKLRPLFEACGEVVOIRPISNRGDFRGYCYVEFK 754

|||||  
Db 310 VLHSSKDSITVFSNLPYSMQEPTKLRPLFACGEVQIRPIFSNRGDFRGYVEFK 369  
QY 755 EKSALQALEMDRKSVEGRPMFVSPVDKSNPDKVFYRSTLEKHKLFISGLPFSC TK 814  
Db 370 EKSALQALEMDRKSVEGRPMFVSPVDKSNPDKVFYRSTLEKHKLFISGLPFSC TK 429  
QY 815 EELEEICKAHGTVKDLRLVTRNAGKPKGLAYVEYENESQASQAVMKMDGMTIKENI KVA 874  
Db 430 EELEEICKAHGTVKDLRLVTRNAGKPKGLAYVEYENESQASQAVMKMDGMTIKENI KVA 489  
QY 875 ISNPPQRKVPKPTRKAPGPMLLPOTYGARGKGRTOQLSLLPRALORPSAAAPQENGP 934  
Db 490 ISNPPQRKVPKPTRKAPGPMLLPOTYGARGKGRTOQLSLLPRALORPSAAAPQENGP 549  
QY 935 AAPAVAAPAAATEAPKMSNADFALFLRK 963  
Db 550 AAPAVAAPAAATEAPKMSNADFALFLRK 578  
RESULT 4  
ABB30113  
ID ABB30113 standard; Peptide; 104 AA.  
XX ABB30113;  
AC  
XX  
XX 01-FEB-2002 (first entry)  
DT  
XX  
XX Peptide #2764 encoded by breast cell single exon nucleic acid probe.  
DE  
XX Human; microarray; single exon probe; gene expression; breast;  
KW disease; cancer.  
KW  
XX  
XX Homo sapiens.  
OS  
XX  
XX W0200157271-A2.  
PN  
XX  
XX 09-AUG-2001.  
PD  
XX  
XX 30-JAN-2001; 2001WO-US00662.  
PF  
XX  
XX 04-FEB-2000; 2000US-0180312.  
PR  
XX 26-MAY-2000; 2000US-0207456.  
PR  
XX 30-JUN-2000; 2000US-0608408.  
PR  
XX 03-AUG-2000; 2000US-0632366.  
PR  
XX 21-SEP-2000; 2000US-0234687.  
PR  
XX 27-SEP-2000; 2000US-0236359.  
PR  
XX 04-OCT-2000; 2000GB-0024263.  
PA (MOLE-) MOLECULAR DYNAMICS INC.  
XX  
XX Penn SG, Hanzel DK, Chen W, Rank DR;  
PI WPI; 2001-496933/54.  
XX  
XX New spatially-addressable set of single exon nucleic acid probes,  
PT useful for measuring gene expression in sample derived from human  
PT breast, comprises number of single exon nucleic acid probes -  
XX  
XX Claim 27; SEQ ID NO 13081; 327pp + sequence listing; English.  
XX  
XX The invention relates to a spatially-addressable set of single exon  
CC nucleic acid probes for measuring gene expression in a sample derived  
CC from human breast and BT 474 cells. The method involves contacting  
CC the probes with a collection of detectably labelled nucleic acids  
CC derived from mRNA of human breast, and then measuring the label  
CC bound to each probe of the microarray. The probes are useful for  
CC verifying the expression of regions of genomic DNA predicted to  
CC encode proteins. They are useful for gene discovery, and for  
CC determining predisposition and/or prognosing breast disease. Gene  
CC expression analysis is useful for assessing the toxicity of chemical  
CC agents on cells. The microarray of this invention presents a far greater  
CC diversity of probes for measuring gene expression, with far less bias

than expressed sequence tag microarrays. The method is suitable for  
CC rapid production of functional information from genomic sequence. The  
CC present sequence is a peptide encoded by a single exon nucleic acid  
CC probe of the invention.  
CC Note: The sequence data for this patent did not form part of the  
CC printed specification, but was obtained in electronic format directly  
CC from WIPO at ftp.wipo.int/pub/published\_pct\_sequences.  
XX  
SQ Sequence 104 AA;  
Query Match 10.5%; Score 526; DB 22; Length 104;  
Best Local Similarity 100.0%; Pred. No. 5.2e-34;  
Matches 104; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 MATAAETSASEPEAESKAGPKADGEDEVKAAATRRKVLRAVAATAATYKTGPAWDQOEE 60  
Db 1 MATAAETSASEPEAESKAGPKADGEDEVKAAATRRKVLRAVAATAATYKTGPAWDQOEE 60  
QY 61 GVSESDGDEYAMASSAESSPGEYEWYDEEEKNQLEIERLEQ 104  
Db 61 GVSESDGDEYAMASSAESSPGEYEWYDEEEKNQLEIERLEQ 104  
RESULT 5  
ABB35284  
ID ABB35284 standard; Peptide; 104 AA.  
XX ABB35284;  
AC  
XX  
XX 04-FEB-2002 (first entry)  
DT  
XX  
XX Peptide #2790 encoded by human foetal liver single exon probe.  
DE  
XX Human; foetal liver; gene expression; single exon nucleic acid probe.  
KW  
XX  
XX Homo sapiens.  
OS  
XX  
XX W0200157277-A2.  
PN  
XX  
XX 09-AUG-2001.  
PD  
XX  
XX 30-JAN-2001; 2001WO-US00669.  
PF  
XX  
XX 04-FEB-2000; 2000US-0180312.  
PR  
XX 26-MAY-2000; 2000US-0207456.  
PR  
XX 30-JUN-2000; 2000US-0608408.  
PR  
XX 03-AUG-2000; 2000US-0632366.  
PR  
XX 21-SEP-2000; 2000US-0234687.  
PR  
XX 27-SEP-2000; 2000US-0236359.  
PR  
XX 04-OCT-2000; 2000GB-0024263.  
PA (MOLE-) MOLECULAR DYNAMICS INC.  
XX  
XX Penn SG, Hanzel DK, Chen W, Rank DR;  
PI WPI; 2001-483447/52.  
XX  
XX Human genome-derived single exon nucleic acid probes useful for  
PT analyzing gene expression in human fetal liver -  
PT  
XX  
XX Claim 27; SEQ ID NO 27919; 639pp + sequence listing; English.  
XX  
XX The invention relates to a single exon nucleic acid probe for  
CC measuring human gene expression in a sample derived from human foetal  
CC liver. The single exon nucleic acid probes may be used for predicting,  
CC measuring and displaying gene expression in samples derived from human  
CC foetal liver. The present sequence is a peptide encoded by a single exon  
CC nucleic acid probe of the invention.  
CC Note: The sequence data for this patent did not form part of the  
CC printed specification, but was obtained in electronic format directly  
CC from WIPO at ftp.wipo.int/pub/published\_pct\_sequences.  
XX  
XX Sequence 104 AA;  
SQ

Qy	1	MATAAETSASEPEAESKAGPKADGEDEVKAAARTRKVLSRAVAATAATYKTMGPAMDQOEE	60
Db	1	MATAAETSASEPEAESKAGPKADGEDEVKAAARTRKVLSRAVAATAATYKTMGPAMDQOEE	60
Qy	61.	GVSESOGDEYAMASSAESPGGEYEWYDEEEKNQLEIRLEEQ	104
Db	61	GVSESOGDEYAMASSAESPGGEYEWYDEEEKNQLEIRLEEQ	104
RESULT 7			
AAID	561114		
ID	AAW561114	standard; Protein; 104 AA.	
XX	XX	AAW561114;	
XX	XX	05-NOV-2001 (first entry)	
DT	XX	Human brain expressed single exon probe encoded protein SEQ ID NO: 28219	
DE	XX	Human; brain expressed exon; gene expression analysis; probe;	
KW	KW	microarray; Alzheimer's disease; multiple sclerosis; schizophrenia;	
KW	KW	epilepsy; cancer.	
XX	XX	Homo sapiens.	
OS	OS	WO200157275-A2.	
PN	PN	09-AUG-2001.	
XX	XX	30-JAN-2001; 2001WO-US00667.	
PR	PR	04-FEB-2000; 2000US-0180312.	
PR	PR	26-MAY-2000; 2000US-0207456.	
PR	PR	30-JUN-2000; 2000US-0608408.	
PR	PR	03-AUG-2000; 2000US-0632366.	
PR	PR	21-SEP-2000; 2000US-0234687.	
PR	PR	27-SEP-2000; 2000US-0236359.	
PR	PR	04-OCT-2000; 2000GB-0024263.	
XX	XX	(MOLE-) MOLECULAR DYNAMICS INC.	
PA	XX	Penn SG, Hanzel DK, Chen W, Rank DR;	
PI	XX	WPI; 2001-483446/52.	
XX	XX	Single exon nucleic acid probes for analyzing gene expression in human	
PT	PT	brains -	
PS	XX	Example 4; SEQ ID NO: 28219; 650pp + Sequence Listing; English:	
XX	XX	The present invention provides a number of single exon nucleic acid	
CC	CC	probes which are derived from genomic sequences expressed in the human	
CC	CC	brain. They can be used to measure gene expression in brain cell samples	
CC	CC	which may enable the diagnosis and improved treatment of nervous system	
CC	CC	diseases such as Alzheimer's disease, multiple sclerosis, schizophrenia,	
CC	CC	epilepsy and cancers. The present sequence is a protein encoded by one o	
CC	CC	the probes of the invention.	
XX	XX	Sequence 104 AA;	
XX	XX	Query Match 10.5%; Score 526; DB 22; Length 104;	
XX	XX	Best Local Similarity 100.0%; Pred. No. 5.2e-34;	
XX	XX	Matches 104; Conservative 0; Mismatches 0; Indels 0; Gaps	
Qy	1	MATAAETSASEPEAESKAGPKADGEDEVKAAARTRKVLSRAVAATAATYKTMGPAMDQOEE	60
Db	1	MATAAETSASEPEAESKAGPKADGEDEVKAAARTRKVLSRAVAATAATYKTMGPAMDQOEE	60
Qy	61	GVSESOGDEYAMASSAESPGGEYEWYDEEEKNQLEIRLEEQ	104
Db	61	GVSESOGDEYAMASSAESPGGEYEWYDEEEKNQLEIRLEEQ	104



```
RESULT 8
AAM68486
ID AAM68486 standard; Protein; 104 AA.
XX
AC AAM68486;
XX
DT 06-NOV-2001 (first entry)
XX
DE Human bone marrow expressed probe encoded protein SEQ ID NO: 28792.
XX
KW Human; bone marrow expressed exon; gene expression analysis; probe;
KW microarray; cancer; leukaemia; lymphoma; myeloma.
XX
OS Homo sapiens.
XX
PN WO200157276-A2.
XX
PD 09-AUG-2001.
XX
PF 30-JAN-2001; 2001WO-US00668.
XX
PR 04-FEB-2000; 2000US-0180312.
XX
PR 26-MAY-2000; 2000US-0207456.
XX
PR 30-JUN-2000; 2000US-0608408.
XX
PR 03-AUG-2000; 2000US-0632366.
XX
PR 21-SEP-2000; 2000US-0234687.
XX
PR 27-SEP-2000; 2000US-0236359.
XX
PR 04-OCT-2000; 2000GB-0024263.
XX
PA (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX
DR WPI; 2001-488901/53.
XX
PT Human genome-derived single exon nucleic acid probes useful for
PT analyzing gene expression in human bone marrow -
XX
PS Example 4; SEQ ID NO: 28792; 658pp + Sequence Listing; English.
XX
XX The present invention provides a number of single exon nucleic acid
XX probes which are derived from genomic sequences expressed in the human
XX bone marrow. They can be used to measure gene expression in bone marrow
XX samples, which may enable the improved diagnosis and treatment of cancers
XX such as lymphoma, leukaemia and myeloma. The present sequence is a
XX protein encoded by one of the probes of the invention.
XX
SQ Sequence 104 AA;
Query Match 10.5%; Score 526; DB 22; Length 104;
Best Local Similarity 100.0%; Pred. No. 5.2e-34;
Matches 104; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MATAAETSASEPEAESKAGPKADGEDEVKAAARTRRKVLRSRAVAATYKTMGPAMDQEE 60
DB 1 MATAAETSASEPEAESKAGPKADGEDEVKAAARTRRKVLRSRAVAATYKTMGPAMDQEE 60
QY 61 GVSESDGDEYAMASSAESPGGEYEWYDEEEKNQLEIERLEE 104
DB 61 GVSESDGDEYAMASSAESPGGEYEWYDEEEKNQLEIERLEE 104
RESULT 9
AAM16294
ID AAM16294 standard; Protein; 104 AA.
XX
AC AAM16294;
XX
DT 12-OCT-2001 (first entry)
XX
DE Peptide #2728 encoded by probe for measuring cervical gene expression.
XX
```

```
KW Probe; human; microarray; gene expression; cervical epithelial cell;
KW cervical cancer.
XX
OS Homo sapiens.
XX
PN WO200157278-A2.
XX
PD 09-AUG-2001.
XX
PF 30-JAN-2001; 2001WO-US00670.
XX
PR 04-FEB-2000; 2000US-0180312.
XX
PR 26-MAY-2000; 2000US-0207456.
XX
PR 30-JUN-2000; 2000US-0608408.
XX
PR 03-AUG-2000; 2000US-0632366.
XX
PR 21-SEP-2000; 2000US-0234687.
XX
PR 27-SEP-2000; 2000US-0236359.
XX
PR 04-OCT-2000; 2000GB-0024263.
XX
PA (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX
DR WPI; 2001-488901/53.
XX
PT Human genome-derived single exon nucleic acid probes useful for
PT analyzing gene expression in human cervical epithelial cells -
XX
PS Claim 27; SEQ ID No 21120; 487pp; English.
XX
XX The present invention relates to human single exon nucleic acid probes
XX (SENP: see AAI10068-AAI28459). The present sequence is a peptide encoded
XX by one such probe. The SENPs are derived from human HeLa cells. The SENPs
XX can be used to produce a single exon microarray, which can be used for
XX measuring human gene expression in a sample derived from human cervical
XX epithelial cells. By measuring gene expression, the probes are therefore
XX useful in grading and/or staging of diseases of the cervix, notably
XX cervical cancer.
XX
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 104 AA;
Query Match 10.5%; Score 526; DB 22; Length 104;
Best Local Similarity 100.0%; Pred. No. 5.2e-34;
Matches 104; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MATAAETSASEPEAESKAGPKADGEDEVKAAARTRRKVLRSRAVAATYKTMGPAMDQEE 60
DB 1 MATAAETSASEPEAESKAGPKADGEDEVKAAARTRRKVLRSRAVAATYKTMGPAMDQEE 60
QY 61 GVSESDGDEYAMASSAESPGGEYEWYDEEEKNQLEIERLEE 104
DB 61 GVSESDGDEYAMASSAESPGGEYEWYDEEEKNQLEIERLEE 104
RESULT 10
AAM28783
ID AAM28783 standard; Protein; 104 AA.
XX
AC AAM28783;
XX
DT 17-OCT-2001 (first entry)
XX
DE Peptide #2820 encoded by probe for measuring placental gene expression.
XX
KW Probe; microarray; human; placenta; antenatal diagnosis;
KW genetic disorder.
XX
OS Homo sapiens.
XX
PN WO200157272-A2.
```



PR 03-AUG-2000; 2000US-0632366.  
 PR 21-SEP-2000; 2000US-234687P.  
 PR 27-SEP-2000; 2000US-236359P.  
 PR 04-OCT-2000; 2000GB-0024263.  
 XX (MOLE-) MOLECULAR DYNAMICS INC.  
 PA Penn SG, Hanzel DK, Chen W, Rank DR;  
 PI WPI; 2002-114183/15.  
 XX  
 DR Spatially-addressable set of single exon nucleic acid probes, used to  
 PT measure gene expression in human lung samples -  
 PT  
 PS Claim 27; SEQ ID NO 27732; 634pp; English.  
 XX  
 CC The invention relates to a spatially-addressable set of single exon  
 CC nucleic acid probes for measuring gene expression in a sample derived  
 CC from human lung comprising single exon nucleic acid probes having one of  
 CC 12614 nucleic acid sequences mentioned in the specification, or their  
 CC complements or the 12387 open reading frames derived from the 12614  
 CC probes. Also included are a microarray comprising the novel set of  
 CC probes; the novel set of probes which hybridise at high stringency to a  
 CC nucleic acid expressed in the human lung; measuring gene expression in a  
 CC sample derived from human lung, comprising (a) contacting the array with  
 CC a collection of detectably labeled nucleic acids derived from human lung  
 CC mRNA, and (b) measuring the label detectably bound to each probe of  
 CC the array; identifying exons in a eukaryotic genome, comprising  
 CC (a) algorithmically predicting at least one exon from genomic sequences  
 CC of the eukaryote; and (b) detecting specific hybridisation of detectably  
 CC labeled nucleic acids from eukaryote lung mRNA, to a single exon probe,  
 CC having a fragment identical to the predicted exon, the probe is included  
 CC in the above mentioned microarray; assigning exons to a single gene,  
 CC comprising (a) identifying exons from genomic sequence by the method  
 CC above and (b) measuring the expression of each of the exons in several  
 CC tissues and/or cell types using hybridisation to a single exon  
 CC microarrays having a probe with the exon, where a common pattern of  
 CC expression of the exons in the tissues and/or cell types indicates that  
 CC the exons should be assigned to a single gene; a peptide comprising one  
 CC of 12011 sequences, mentioned in the specification, or encoded by the  
 CC probes/open reading frames (ORF). The probes are used for gene  
 CC expression analysis, and for identifying exons in a gene, particularly  
 CC using human lung derived mRNA and for the study of lung diseases  
 CC such as asthma, lung cancer, chronic obstructive pulmonary disease  
 CC (COPD), interstitial lung disease (ILD), familial idiopathic pulmonary  
 CC fibrosis, neurofibromatosis, tuberous sclerosis, Gaucher's disease,  
 CC Niemann-Pick disease, Hermansky-Pudlak syndrome, sarcoidosis, pulmonary  
 CC haemosiderosis, pulmonary histiocytosis, lymphangioleiomyomatosis,  
 CC pulmonary alveolar proteinosis, Karagener syndrome, fibrocystic  
 CC pulmonary dysplasia, primary ciliary dyskinesia, pulmonary hypertension  
 CC and hyaline membrane disease. The present sequence is a peptide/protein  
 CC encoded by a single exon probe of the invention.  
 CC Note: The sequence data for this patent did not form part  
 CC of the printed specification, but was obtained in electronic  
 CC format directly from WIPO at  
 CC ftp.wipo.int/pub/published\_pct\_sequences.  
 XX  
 SQ Sequence 104 AA;  
 Query Match 10.5%; Score 526; DB 23; Length 104;  
 Best Local Similarity 100.0%; Pred. No. 5.2e-34;  
 Matches 104; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 MATAETASAEPEAESKAGPKADGEDEVKARTRKVLSRAVAATYKTMGPANDQEE 60  
 |||||||  
 Db 1 MATAETASAEPEAESKAGPKADGEDEVKARTRKVLSRAVAATYKTMGPANDQEE 60  
 |||||||  
 QY 61 GVSSEGDGEYAMASSAESSPGEYEWYDEEEKNLEIRLEEQ 104  
 |||||||  
 Db 61 GVSSEGDGEYAMASSAESSPGEYEWYDEEEKNLEIRLEEQ 104  
 |||||||

RESULT 13

CC source of primers and probes, in the detection of ORFX genomic sequences  
 CC or transcripts, in the identification and cloning of homologous  
 CC sequences, in genetic diagnosis, and in forensic biology. The ORFX  
 CC nucleic acids may additionally be used to produce transgenic animals  
 CC which may be useful for studying the function and/or activity of ORFX  
 CC protein, and in drug screening. The ORFX proteins may also be used as  
 CC immunogens to generate specific antibodies, which are useful in the  
 CC diagnosis, treatment and monitoring of ORFX-associated diseases.  
 XX SQ Sequence 108 AA;

Query Match 10.3%; Score 513; DB 23; Length 108;  
 Best Local Similarity 87.0%; Pred. No. 6e-33;  
 Matches 94; Conservative 9; Mismatches 5; Indels 0; Gaps 0;  
 QY 367 VHNRAIRNCPWTVVALWSRYLLAMERHGVHDHVISVTFEKKALNAGFTQATDYVEIQAYLD 426  
 DB 1 VHSRAVRNCPWTVVALWSRYLLAMERHGLDHTISATFNALSAGFIQATDYVEIQAYLD 60  
 QY 427 YLRRRVDFKQDSSKELEELRAAFTRALEYLKOVEEERFNESEGDPPSCVI 474  
 DB 61 YLRRRVDFKQDSSKELEELRSMFTRALEYLKOVEEERFSESGDPSCLI 108

RESULT 14  
 ID ABB59801 standard; Protein; 941 AA.

AC ABB59801;

DT 26-MAR-2002 (first entry)

DE Drosophila melanogaster polypeptide SEQ ID NO 6195.

KW Drosophila; developmental biology; cell signalling; insecticide;  
 KW Pharmaceutical.

OS Drosophila melanogaster.

XX WO200171042-A2.

PD 27-SEP-2001.

XX 23-MAR-2001; 2001WO-US09231.

XX 23-MAR-2000; 2000US-191637P.

PR 11-JUL-2000; 2000US-0614150.

XX (PEKE ) PE CORP NY.

XX Venter JC, Adams M, Li PWD, Myers EW;

XX WPI; 2001-656860/75.

XX N-PSDB; ABL03904.

PT New isolated nucleic acid detection reagent for detecting 1000 or more  
 PT genes from Drosophila and for elucidating cell signalling and cell-cell  
 PT interactions

XX Disclosure; SEQ ID.NO 6195; 2lpp + Sequence Listing; English.

CC The invention relates to an isolated nucleic acid detection reagent  
 CC capable of detecting 1000 or more genes from Drosophila. The invention is  
 CC useful in developmental biology and in elucidating cell signalling and  
 CC cell-cell interactions in higher eukaryotes for the development of  
 CC insecticides, therapeutics and pharmaceutical drugs. The invention  
 CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA  
 CC sequences (ABL01840-ABL16175) and the encoded proteins  
 CC (ABBS57737-ABBS72072).

CC The sequence data for this patent did not form part of the printed  
 CC specification, but was obtained in electronic format directly from WIPO  
 CC at ftp.wipo.int/pub/published\_pct\_sequences.

XX

SQ Sequence 941 AA;  
 Query Match 8.5%; Score 425; DB 22; Length 941;  
 Best Local Similarity 20.3%; Pred. No. 1.7e-24;  
 Matches 221; Conservative 151; Mismatches 347; Indels 372; Gaps 40;  
 QY 7 TSAPAEABSKAGPKADGEDEYKAAARTRRKVLRSRAAAATYKTMGPAMDQOQEGYSED 66  
 DB 87 SSDDEPSVEETEGGNAAGR-----GRARNDSSSSDDVGVIE--GSELSNNESSDSD 138  
 QY 67 GDEYAMASSAESPGYEYWEYDEEEKNOLETERLEEQLSINYDYNCHVDLIRLRLEG 126  
 DB 139 SD-----SDNAGGNNOLERSY---QELNALPSKKFAQMV-----LIGIAFKLN 179  
 QY 127 ELTKVRMARQKSEIFPLTEELWLEWLDHEISMAQDGLDREHYVDLFEKAVDKYICPNIW 186  
 DB 180 DIEKTESSVLEQLNATVPAHVWLKYLKARLVVTTQDDEERKAFEEQCAKALGYYS---- 235  
 QY 187 LEYGQYSGVIGIGQGGLEKRVSRFVERALS SVGLHMTKGLALWEAYREFESAIVEAARLEK 246  
 DB 236 IPLSEYVNVYLVLDQGNVQN-----HV-----LW-----AKLLA 263  
 QY 247 VHSLFRQLAIPLYDMEATFAEYEE-----WS-----EDPIPESVIYQNYKA 288  
 DB 264 DYDVERPDGDKLRSLISTITDENEAAAFVEMLOKHCVTWTNCVEORQMIKSVYDKFKQH 323  
 QY 289 LOOLEKYKPYEEALLQAEAPRLAEYQAYI--DPE-MKIGDP---ARTOLIFERALVENCVLV 343  
 DB 324 LDETRQWDWE-----QKHAHYVDVETLSLDDDKNAVIRFIFERSVAKEPIV 372  
 QY 344 PDLWIRYSQYLD-----RQLK-----VKDLVLSVHNRAIRNCP 376  
 DB 373 DVLWLSYIEFIOFEGVTVPENEDENEVTAEMVAKRAKRLGKGLRNLTDLANRGVRSHP 432  
 QY 377 WTVALMSRYLLAMERHGVHDHVISVTFEKKALNAGFTQ--ATDYVEIWAQYLDLRRRV-D 433  
 DB 433 -SVQLNHRFLDLMERS--DFELAEVDEEIRL---ILQRIVTDMDTVLHLDYLAIRN 486  
 QY 434 FKQDSSKELEELRAAFTRALEYLKOVEEERFNESEGDPPSCVIMONWARIARLNNMQAR 493  
 DB 487 TNASDEQQVVASLRAAFNHAWE-----ELTVLYGDAQDTRYEVQLQMAQVEYTLQSGPDNGR 542  
 QY 494 ELWDSIM-TRGNAKYANMLEYNLER-----AHGDTQHCRKALHRAVQCTSDYPEHVCEV 548  
 DB 543 EIWRQIMGYPGSGSIRGLLMLNFAQMESEYNGHGTRDVLRLKALSQPV---LENGLMVQEF 599  
 QY 549 LLTMERTESLED-----WDIAVQKTETRLARVNEQRMAKAEKALVQOEEKEAEQRK 602  
 DB 600 FRRYERCYGTYESIAACQALDLPVEYVKPR-SRIKFNQSQSAVPROOKLAPROOQOQTNRE 658  
 QY 603 RARAEKKALKKKKIRGPEKRGADDEDEKEMGDEDEEQPSKRRRRVENSIPAAGETQNV 662  
 DB 659 PLNREQ-----RRQHEQOQOQ---OQOQKHGIIKSRTE---PSGAT----- 696  
 QY 663 AAGPAGKCAADVPEPPSKQKKAASLKRDMPKVLHDSKDSITVFVSNLPYSMOEPDTKL 722  
 DB 697 -----SPFSKVK----- 703  
 QY 723 RPLFEACGEVQIRPIFSNRGDPGRGYCYVEFKKEKSALEMDRKSVEGRPMFVSPCVD 782  
 DB 704 -----GPANAEAKESN----- 714  
 QY 783 KSKNPDKFVRYSTSLKHKLFISGLPFSCCTKEELEICKAHGTVCKDLRLVTNRACKPKG 842  
 DB 715 -----FKYSPNMEIKFVNRNHPACSKSELHELFSFGTINKDVLVHLKLNKQFKG 765  
 QY 843 LAYVEYENESQASQAVKMDGMTIKENIIKVAISNPPQRKVPE-KPETRAKAPGPMLLPQ 901  
 DB 766 IAYVEEKEPCEAQRVAGRGCLFKCMNISVALSNPPRPTSAVKPSV-AP----- 815  
 QY 902 TYGARGKGTQSLPLRALQRPSSAAAPQ-----AENGPAAPAAVAAPAT----- 946  
 DB 816 -----KRRVPTSLIPTTLVRQEAACKLRLLLPEPGDISSTSSASVDVAIKREANGEOK 869



Result No.	Query			ID	Description
	Score	Match	Length		
1	272	5.4	414	1	US-07-667-276A-4
2	224.5	4.5	714	2	US-08-990-114-3
3	224.5	4.5	714	4	US-09-241-333-3
4	204	4.1	475	2	US-08-698-407-4
5	204	4.1	475	3	US-09-195-855-4
6	202.5	4.1	1886	4	US-08-938-105-3
7	201	4.0	485	1	US-07-881-075-1
8	201	4.0	485	1	US-08-120-827-1
9	201	4.0	485	1	US-08-478-675-1
10	185.5	3.7	976	4	US-09-104-324B-4
11	185.5	3.7	1939	4	US-09-310-187A-1
12	181	3.6	255	4	US-09-370-838-183
13	178.5	3.6	323	1	US-07-667-276A-7
14	173.5	3.5	444	1	US-07-881-075-3
15	173.5	3.5	444	1	US-08-120-827-3
16	173.5	3.5	444	1	US-08-478-675-3
17	170	3.4	380	1	US-07-881-075-51
18	170	3.4	380	1	US-08-120-827-51
19	170	3.4	380	1	US-08-478-675-51
20	169.5	3.4	759	1	US-08-676-967-1
21	169.5	3.4	759	1	US-08-676-974-1
22	169.5	3.4	759	2	US-09-098-487-1
23	168.5	3.4	1388	2	US-08-685-576-1
24	168	3.4	652	4	US-09-347-833-2
25	166.5	3.3	2101	1	US-08-466-390-4
26	166.5	3.3	2101	1	US-08-470-950-4
27	166.5	3.3	2101	1	US-08-467-781-4





SEQUENCE CHARACTERISTICS:  
LENGTH: 714 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
IMMEDIATE SOURCE:  
LIBRARY: GenBank  
CLONE: 128842  
US-09-241-333-3

Query Match 4.5%; Score 224.5; DB 4; Length 714;  
Best Local Similarity 23.5%; Pred. No. 5.8e-09;  
Matches 94; Conservative 67; Mismatches 134; Indels 105; Gaps 16;  
QY 583 AAEKEAALVQOEIEKAEKQKRAAE-----KKALKKKKIRGPEKRGAEDEDE 630  
DB 187 ASEDEDEDEDEDEDEDEDEEAMBITPAKGGKAKAPKVPVAKNVAEEDDDDE 246  
QY 631 KEMGDDEEQ-----PSKRRR---VENSIPAAGETQNV----- 661  
DB 247 EDEDEDEDEDEDEDEDEEPEVPKAPGKRRKKEWTKQKEVPEA-KKQKVGSEST 305  
QY 662 -----VAAGPACK--CAAVDVPEPSKQK-----EKAA 686  
DB 306 TPFNLFIGNLPNKSVAELKVAISPEFAKNDLAVVDVGTNRKFGYVDFESAEDLEKAL 365  
QY 687 SL-----KRDMPKVLHDSSK--DSITVFVSNLPSYMOEPDTKLRPLFEACGEVVO 734  
DB 366 ELTGKLVGNEIKLEKPKG-RDSKKVRAARTLLAKNLSFNTEDE--LKEVPE---DALE 419  
QY 735 IRPTPSNRGDFRGYVFEKESKALQALEMDR-KSVEGRPMFVSPVDKSNPDKFYR 793  
DB 420 IR-LVSDQKSGKIAYIEFKSEADAENLEEKQAEIDGRSVLYTGEKGQORBTCKN 478  
QY 794 YSTLEKKHLFISGLPFSCTEKEELEICKAHCTYKDLRLVNRACKPKGLAYVEVENESQ 853  
DB 479 STWGESKTLVLSNLSYSATEETLQEFVEKATFIK---VPONQQKSGYAFIEFASPED 535  
QY 854 ASQAVKMKDGMTIKENIKVAISNPPQKVPKPTRKAP 893  
DB 536 AKEALNSCKMEIEGRTIRLELOGP-----RGSPNARSQP 570

RESULT 4  
US-08-698-407-4  
Sequence 4, Application US/08698407  
Patent No. 5856128  
GENERAL INFORMATION:  
APPLICANT: Bandman, Olga  
APPLICANT: Au-Young, Janice  
APPLICANT: Hawkins, Phillip R.  
APPLICANT: Hillman, Jennifer L.  
TITLE OF INVENTION: A NOVEL HUMAN NUCLEIC ACID BINDING PROTEIN  
NUMBER OF SEQUENCES: 5  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Incyte Pharmaceuticals, Inc.  
STREET: 3174 Porter Drive  
CITY: Palo Alto  
STATE: CA  
COUNTRY: U.S.  
ZIP: 94304  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: DOS  
SOFTWARE: FastSeq Version 1.5  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/698,407  
FILING DATE: Filed Herewith  
ATTORNEY/AGENT INFORMATION:  
NAME: Billings, Lucy J.  
REGISTRATION NUMBER: 36,749  
REFERENCE/DOCKET NUMBER: PF-0116 US

TELECOMMUNICATION INFORMATION:  
TELEPHONE: 415-855-0555  
TELEFAX: 415-845-4166  
INFORMATION FOR SEQ ID NO: 4:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 475 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
IMMEDIATE SOURCE:  
LIBRARY: GenBank  
CLONE: 267188  
US-08-698-407-4

Query Match 4.1%; Score 204; DB 2; Length 475;  
Best Local Similarity 22.9%; Pred. No. 1.2e-07;  
Matches 85; Conservative 68; Mismatches 174; Indels 44; Gaps 12;  
QY 559 LEDMD-IAVQKTETRLARVNEQRMKAEEKAAALVQOEIEKAEQKQKRAAEKALKKKKK 616  
DB 1 MSDDFEFERQLNENKQERDKENHRKRSRSDRKRKRSRDRRNRDQSRASDRRR 60  
QY 617 IRGPKRGAEDEDEKEMGDDEEQSKRRRVENSIPAAE--TONVEAAGPAGCAAVD 674  
DB 61 RSKPLTRGAKEEHGLIRSPRHEKKKKVKRYWDPVPPGFHEITPMQYKAMQAAGQIPATA 120  
QY 675 VEPSPKQEKAAKSLKRDMPKVLHDSSKDSITVFVSNLPSYMOE-----PDTKLRLPLF 726  
DB 121 L-LPTMTDGLAVTTPVPVGSQMTROARLLYGVNIPFGITEEAMMDFFNAQMLGGLT 179  
QY 727 EACGE--VVQIRPFRGDFRGYVFEKESKALQALEMDRKSVEG-----RPMFV 777  
DB 180 QAPGNPVLAVQI-----NQDKNFALEFRSVDETTQAMAFDGIIFQOSLKIRPHDY 232  
QY 778 SPCVDKSNPDKVPYRSTSL--EKHKLFIISGLPFSCTEKEELEICKAHGVKDLRLVT 834  
DB 233 QPLPKMSENPSYVVPVSVTPVPSAHKLFIFGLPLNYLNDQVKELLTSFGPLKAFNLVK 292  
QY 835 NRA-CKPKGLAYVEVENESQASQAVKMKDGMTIKENIKV-----AISNPPQKRV 883  
DB 293 DSATGLSGYAFCEYVDINVTQATAGLMQGLKLLVQRASVGAKNATLVSPPS-TI 351  
QY 884 PEKPTRKAPG 894  
DB 352 NQTPVTLQVPG 362

RESULT 5  
US-09-195-855-4  
Sequence 4, Application US/09195855  
Patent No. 6015788  
GENERAL INFORMATION:  
APPLICANT: Bandman, Olga  
APPLICANT: Au-Young, Janice  
APPLICANT: Hawkins, Phillip R.  
APPLICANT: Hillman, Jennifer L.  
TITLE OF INVENTION: A NOVEL HUMAN NUCLEIC ACID BINDING PROTEIN  
NUMBER OF SEQUENCES: 5  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Incyte Pharmaceuticals, Inc.  
STREET: 3174 Porter Drive  
CITY: Palo Alto  
STATE: CA  
COUNTRY: U.S.  
ZIP: 94304  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: DOS  
SOFTWARE: FastSeq Version 1.5  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/195,855

FILING DATE:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/698,407  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: Billings, Lucy J.  
REGISTRATION NUMBER: 36,749  
REFERENCE/DOCKET NUMBER: PF-0116 US  
TELEPHONE: 415-855-0555  
TELEFAX: 415-845-4166  
INFORMATION FOR SEQ ID NO: 4:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 475 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
IMMEDIATE SOURCE:  
LIBRARY: GenBank  
CLONE: 267188  
US-09-195-855-4

Query Match 4.1%; Score 204; DB 3; Length 475;  
Best Local Similarity 22.9%; Pred. No. 1.2e-07;  
Matches 85; Conservative 68; Mismatches 174; Indels 44; Gaps 12;  
QY 559 LEDWD-IAYQVETRLARVNEORMKAAEKAALVQEEBKAQR-KRARAEEKALKKKK 616  
DB 1 MSDDFEFQNLNENQERDKENHRKRSRDRKRKRSDRRNRDQSRSDRRR 60  
QY 617 IRGPKRGADDEKWDGDEEOPSKRRRVENSIPAAE--TQNVVEAAGPAGCAAYD 674  
DB 61 RSKPLTRGAKKEHGGGLIRSPREKKKKYKWDVPPGPEHITPMQYKAMQAAGQIPATA 120  
QY 675 VEPSPQKEKAASLRDKPKVLHDSKDSITVFSNLPYSMQE-----PDTKLRLPLF 726  
DB 121 L-LPTMTDGLAVTPTFPVVGVSQMTQARRLYVGNIPGITEAMMDFNMQMRIGGIT 179  
QY 727 EACGE---VVOIRPFSNRGFRGYCYVEFKESKALQALEMDRKSVEG-----RPMFV 777  
DB 180 QAPGNVLAVQI-----NODKNFALEFRSVDETTQAMAFDGIIFQOQSLKIRPHDY 232  
QY 778 SPVCKSNKNDPKVRYSTSL---EKHLFISGLPFSCTKEBEIEICKAHGTVDKDLRLVT 834  
DB 233 QPLPGMSENPVYVPGVSTVVPDPSAHKLFIGGLPNYNDQVKELLTTSFGPLKAFNLVK 292  
QY 835 NRA-GKPKGLAYVENESQASQAVMKMDGMTIKENIKV-----AISNPPQKV 883  
DB 293 DSATGLSKGYAFCEYVDINVTQATAGLNGMLQDKKLLVQRASVGAKNATLVSPPS-TI 351  
QY 884 PEKPTRKAPG 894  
DB 352 NOTPTVLQVP 362

RESULT 6.  
US-938-105-3  
Sequence 3, Application US/08938105  
Patent No. 6353151  
GENERAL INFORMATION:  
APPLICANT: Leinwand, Leslie A.  
APPLICANT: Vikstrom, Karen L.  
TITLE OF INVENTION: TRANSGENIC MODEL FOR HEART FAILURE  
NUMBER OF SEQUENCES: 3  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Sheridan Ross P.C.  
STREET: 1700 Lincoln St., Suite 3500  
CITY: Denver  
STATE: CO  
COUNTRY: U.S.A.  
ZIP: 80203  
COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/938,105  
FILING DATE:  
CLASSIFICATION:  
ATTORNEY/AGENT INFORMATION:  
NAME: Crook, Wanneil M.  
REGISTRATION NUMBER: 31,071  
REFERENCE/DOCKET NUMBER: 3595-4  
TELEPHONE: (303) 863-9700  
TELEFAX: (303) 863-0223  
INFORMATION FOR SEQ ID NO: 3:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1886 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-938-105-3  
Query Match 4.1%; Score 202.5; DB 4; Length 1886;  
Best Local Similarity 19.6%; Pred. No. 1.2e-06;  
Matches 195; Conservative 154; Mismatches 342; Indels 303; Gaps 46;  
QY 16 SKAGPKADGEDEVAARTRRKVLRAVAATAATKTMGPADWQOEGVSESDGDVAMASS 75  
DB 970 TSKVKLEQVDDLSGSEQEKVKRMDLERAKKLEGDL-KUTOSIMDLENDKKLQLEBK 1028  
QY 76 AESSPEGYEWYDEEEKNQLEIE-----RLBEQLSINVYDYNCHVDLIRLLRLEGELTK 130  
DB 1029 LKKK-----EPDISQNSKIEDEQALALQLOKKLEN-----QARIEELELEA 1073  
QY 131 VMAKQKSEI-FPPIELWLEWLHDEISMAODGLDREHYVDLFEKAVKDYICPNILWEY 189  
DB 1074 ERTARAKVEKLSDTREL-----EISERLEA----- 1102  
QY 190 GOYSVGGIGQKGLKVRVFERALSSVGLHMTKGLALWEAYREFESAIVEAARLEKVHS 249  
DB 1103 -----GG-----ATSVQIENK-----KRAEFQK----- 1122  
QY 250 LFEROLAIPLYDMEATFAEYEMSEDPIPE--SVIONYNKALOQLEKYK----- 296  
DB 1123 -MRDLEELQHEATAAARLKHADSVAEELGEQIDNLRVQKLEKSEKFEKLEDDVT 1181  
QY 297 PYEBALLQAE-----PRLAEYQAYIDPEMKI-----GDPARTQ 330  
DB 1182 SHMEQIIRKAKANLEKVSRTLEDQAN-EYRVKLEEAQRSINDFTTORAKLTQENGELAR-Q 1239  
QY 331 LIFERALVENCULVPDLW-----IRYSOYLD---RQL-----KVQDLVLSVHNRAIRNCWP 377  
DB 1240 LEEKEALI-----WQTRGKLSYTOQMEDLKQLEEGKAKNALAHALQSARHDCD- 1290  
QY 378 TVALMSRYLLAMERHGVHDQVIVSVTFEAKNALNAGFIQATDYVEIWOAYLDYLRVRVDFKOD 437  
DB 1291 --LLREQYEEMEAELQRVLS-----KANSEVAQW-----RTKYETD 1327  
QY 438 SSKLEELRAAFTRALEYLVKQVEERFNEBSPGVIMONWARIARICNNMQ----- 490  
DB 1328 AIQRTTEELEEAKKLAQRL-QDAEEAV-EAVNAKCSLE---KTHRLQNEIEDLMVDVVE 1382  
QY 491 -----KARELWDSIMTRGNAYKAMMWEYNLERAHGDTQHCRAKALHRAVQCTSD 540  
DB 1383 RSNAALADKKQRNFDKILAEWKQKYESQSE---LESSQKEARSLTELEKLNAYEE 1439  
QY 541 YPEHVCEVLLTMRTEGSLED--WDIAVQ-----KTETRLARVNEQ-RMKAAEKAALVQ 592  
DB 1440 SLEH-----LETFKRENKLQEEISDLTQLEGGGKNVHELEKIRKQLEVEKLEQSALEE 1495  
QY 593 QEEE-KAEQRKRAEAKKALKKKKIRGPEKGADEDEKENG-----DD 636

Db 1496 AEASLEHBEKGILRAQLEFNOIKAEI-----ERKLAEKDEMEQAKRNHLRVVDSLQTSIDA 1552  
QY 637 EEEQPSKRRRYVENSIPAAGETONVEAAGPAGKCAADVDEPPSKOKEKAAKSLKRDMPKV- 695  
Db 1553 ETRSNEALRYKKM--EGDLNEMEIQLSQANRTAS-----EAKHLKNAQAHLKDTQLD 1607  
QY 696 ----LHSSKDSITVFSNLPYSMOEPTKRLPLFEAGGEVYQIRPFSNRGDFRGYCV 751  
Db 1608 DAVRANDLKENIAI-----VERRNTLLQ-----AELELRVVEQTEQSRKLAEQ 1653  
QY 752 EFKEKSALQALEMDRKSVEGRPMFVSPCVDSKNPDKVFRYSTSL-----EK 800  
Db 1654 ELIETSERVQLHLQSNTSL-----INOKKMDADLSQLQTEVEEAVOEACRANAEK 1703  
QY 801 HKLTSGLPFSCTKEELEECIAH-----GTVKDLRLVTNRA-----GPKGLAY 845  
Db 1704 AKKAITDAMMAEELKREODTSAHLERMKKNMEQTIKDLQHLDEAEQIALKGGKKLOK 1763  
QY 846 V-----EYENESQASQA--VMKMDGMTIKENIK 872  
Db 1764 LEARVRELENELEAEQKRNAESVKGMRKSERRIK 1797

## RESULT 7

US-07-881-075-1  
; Sequence 1, Application US/07881075  
; Patent No. 5444149

; GENERAL INFORMATION:  
; APPLICANT: KEENE, JACK D.  
; APPLICANT: KING, PETER H.  
; APPLICANT: LEVINE, TODD  
; TITLE OF INVENTION: METHODS AND COMPOSITIONS USEFUL IN THE  
; TITLE OF INVENTION: RECOGNITION, BINDING AND EXPRESSION OF RIBONUCLEIC ACIDS  
; TITLE OF INVENTION: INVOLVED IN CELL GROWTH, NEOPLASIA AND IMMUNOREGULATION  
; NUMBER OF SEQUENCES: 51  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: OBLON, SPIVAK, MCCLELLAND, MAIER & NEUSTADT,  
; STREET: 1755 Jefferson Davis Highway, Fourth Floor  
; CITY: Arlington  
; STATE: Virginia  
; COUNTRY: U.S.A.  
; ZIP: 22202

; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/07/881,075  
; FILING DATE: 19920511  
; CLASSIFICATION: 530  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Oblon, No. 5444149man F.  
; REGISTRATION NUMBER: 24,618  
; REFERENCE/DOCKET NUMBER: 714-154-0  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (703)521-4500  
; TELEFAX: (703)486-2347  
; TELEX: 248855 OPAT UR  
; INFORMATION FOR SEQ ID NO: 1:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 485 amino acids  
; TYPE: AMINO ACID  
; TOPOLOGY: unknown  
; MOLECULE TYPE: protein

US-07-881-075-1

Query Match 4.0%; Score 201; DB 1; Length 485;

Best Local Similarity 20.7%; Pred. No. 2.1e-07;  
Matches 91; Conservative 73; Mismatches 136; Indels 140; Gaps 16;

QY 564 IAVQKTETRLA-----RVNEQRMKAAEKAALVQOEEKEAEQKRRAEKALKKK 614

Db 64 LQVQOQOQTQAVAAAAAAVTQLOQOQAVVAQVQOQOQAAAAVVAQVQOQAVVQ 123  
QY 615 KKIRCEPKRGADEDEKEKEMGDEEBEOPSKRRRRENSIPAAGETONVEAAGPAGKCAAVD 674  
Db 124 POQAQPTNG-----NAGSGSQN-----GSG----- 145  
QY 675 VEPFSKQKEKAASLRKDPKVLHDSKSDSITVFSNLPYSMOEPTKRLPLFEAGGEVYQ 734  
Db 146 -----STETRTNLIVNLPQTWTEDE--IRSLFSSVGEIES 179  
QY 735 IR-----PIFSNRGDRPGYCYVEFEKEKSALQA-----LEMDRKSVEGRP 774  
Db 180 VKLIRKDSQVYIDPLNPOAPSKGSLGXGVYVYRPODAEQAVNVLNGLRLQNKTI--KV 237  
QY 775 MFVSPCVDSKNPDKVFRYSTSLKHLFTSGLPFSCTKELEICRAHGTVKDLRLVT 834  
Db 238 SFARSSDAIKAN-----LYVSGLPKTMQTQOEELEAIFAPFGAITSRLQ 283  
QY 835 NRAG--KPKGLAYVEYENESQASQAVMKMDGMTIKE--NIKVAISNPP-----QRKV 883  
Db 284 NAGNDTQTKGVCFIRFDKREERATRAIIALNGTTPSSCTDPIVVKFSNTPGSTSKIQPOL 343  
QY 884 PE--KPTRKAPGGPMLLPQTVGARGKGTQLS-----LLPRALQRPSSAAAPQARN 932  
Db 344 PAFLNPQLVRRIGGAMHTPV-----NKGLARESPMAGDMLDVMPLNGLCAAAAAATTLAS 398  
QY 933 GPAAAPAV-----AAPATEA 948  
Db 399 GPGGAYPIFYVNLAPETEEA 418

## RESULT 8

US-08-120-827-1  
; Sequence 1, Application US/08120827  
; Patent No. 5525495

; GENERAL INFORMATION:  
; APPLICANT: KEENE, JACK D.  
; APPLICANT: KING, PETER H.  
; APPLICANT: LEVINE, TODD  
; TITLE OF INVENTION: METHODS AND COMPOSITIONS USEFUL IN THE  
; TITLE OF INVENTION: RECOGNITION, BINDING AND EXPRESSION OF RIBONUCLEIC ACIDS  
; TITLE OF INVENTION: INVOLVED IN CELL GROWTH, NEOPLASIA AND IMMUNOREGULATION  
; NUMBER OF SEQUENCES: 101  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: OBLON, SPIVAK, MCCLELLAND, MAIER & NEUSTADT,  
; STREET: 1755 Jefferson Davis Highway, Fourth Floor  
; CITY: Arlington  
; STATE: Virginia  
; COUNTRY: U.S.A.  
; ZIP: 22202

; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/120,827  
; FILING DATE: 15-SEP-1993  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Oblon, No. 5525495man F.  
; REGISTRATION NUMBER: 24,618  
; REFERENCE/DOCKET NUMBER: 714-158-0 CIP  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (703)413-3000  
; TELEFAX: (703)413-2220  
; TELEX: 248855 OPAT UR  
; INFORMATION FOR SEQ ID NO: 1:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 485 amino acids  
; TYPE: amino acid



;; SOFTWARE: Wordperfect  
;; CURRENT APPLICATION DATA:  
;; APPLICATION NUMBER: US/09/104.324B  
;; FILING DATE: 25-June-1998  
;; CLASSIFICATION: 435  
;; PRIOR APPLICATION DATA:  
;; APPLICATION NUMBER: 08/892,702  
;; FILING DATE: 15-July-1997  
;; ATTORNEY/AGENT INFORMATION:  
;; NAME: Hanson, No. 6232460man D.  
;; REGISTRATION NUMBER: 30,946  
;; REFERENCE/DOCKET NUMBER: LUD 5491  
;; TELECOMMUNICATION INFORMATION:  
;; TELEPHONE: (212) 318-3000  
;; TELEFAX: (212) 752-5958  
;; INFORMATION FOR SEQ ID NO: 4:  
;; SEQUENCE CHARACTERISTICS:  
;; LENGTH: 976 amino acids  
;; TYPE: amino acid  
;; TOPOLOGY: linear  
US-09-104-324B-4

Query Match 3.7%; Score 185.5; DB 4; Length 976;  
Best Local Similarity 19.5%; Pred. No. 9.5e-06;  
Matches 189; Conservative 152; Mismatches 351; Indels 277; Gaps 43;  
QY 28 EVKAARTRRKVLRAVAATYKMGPAWDOQE-----EGVSESDGEYAMASSAES 78  
DB 175 ENNATHILCNLTKETCARSAEKTKEYEETROQVYMDLNNIEKMITAFQELRVQAE 234  
QY 79 SPGEYEWYDEEEKNOLEIEERLEQSLINVDYNCHVDLRLRLRLEGELEKVMAROKM 138  
DB 235 SRLEMHFKLDEYK-----IQHLEQYKKEINDREKQVSL-----LIQITEKENKM 282  
QY 139 SEIPLPLTELWLEWLBDSMAQD--GLDRHVDLFEK---AVKDYICPNWLEYGQYS 193  
DB 283 KDLTFLEE-----SRDVNOLKEKTKLQSENKQSIKQHLTK----- 323  
QY 194 VGGIGQGGLEKVRSVFERASS-----VGLHMTKGLALWEAYR----- 232  
DB 324 -----LEDIKVSLQSRVSTOKALEEDIQATKTKTCQTEKETQMEESNARAHAHS 374  
QY 233 -----EFESAI-----VEAARLEKHSFRRLQALPLYDMEATFAEYEWSESDPIPES 280  
DB 375 FVWTEFTVCSLELLRTEQORLEKNE-----QLKILTMELQKSSLEEMTK----- 424  
QY 281 VIONYNKALQOLEKVKPEEALL--QAEPRLAEYQVADIFEMKIGDIPARTQILIFERALV 338  
DB 425 LTNKVELEELKVLGKETLLYENKQFEKIAE-----ELKGTQELIGLLQAREKE 477  
QY 339 ENCLVPDLIRYS-----QYLDROLKVKDLVLSVHNRAIRNCPWTVALWSRYLLAMERH 392  
DB 478 -----VHDEIOLITAITSEQYYSK--VKDLKTELENEKLNKTELTSHC---NKLSLENK 528  
QY 393 GVDHQVIVTPEKALNAGFIQATQVETWQAYLDYLRVRVDFKODSSKELEELRAAFTRA 452  
DB 529 ELTQETSDMTLE-----LKNQOEDINNKKQEEERMLQIENLOETETQLRNE 575  
QY 453 LEYLKQVEEERFNSGDPSCVIMQWARIERL-----CNNMOKARELWDSIMTRGNK 506  
DB 576 LEYVREELKQRDE-----VKCKLDKSEENCNLRK-----QVE 609  
QY 507 YANMWLEYNLERAHGDTQHRKALHRAVOCTSD---YPEHVCEVLLTMTERTGSLSDW 562  
DB 610 NKNYIE-----ELQENKALKKGTAEKQNLNVETKVNKLELELSAKQKFGEI 660  
QY 563 DIAVQKTTRLARVNEORM-----KAAEKEAALVQOEEKAEQKREARAEKALKKKK 615  
DB 661 TDYQK-BIEDKTISEENLLEVEKAKVIADAEVAKLQEKIDKRCQH--IAEWALMEKH 717  
QY 616 KIRGPEKGADEDEKENG-----DDEEOPSKRRRVENSIPAGETQNVVEAAGPAGCAA 672  
DB 718 KHOYDK-----IIEERSELGLYKSKBQEQSSLRASLEI-----ELSNLKAELLSYKKOLE 768

RESULT 11  
US-09-310-187A-1  
; Sequence 1, Application US/09310187A  
; Patent No. 6358751  
; GENERAL INFORMATION:  
; APPLICANT: Benichou, Gilles  
; APPLICANT: Fedoseyeva, Eugenia  
; TITLE OF INVENTION: Involvement of Autoantigens in Cardiac  
; TITLE OF INVENTION: Graft Rejection  
; FILE REFERENCE: UCSF-090  
; CURRENT APPLICATION NUMBER: US/09/310,187A  
; CURRENT FILING DATE: 1999-05-12  
; NUMBER OF SEQ ID NOS: 3  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 1  
; LENGTH: 1939  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-310-187A-1

Query Match 3.7%; Score 185.5; DB 4; Length 1939;  
Best Local Similarity 19.6%; Pred. No. 2.6e-05;  
Matches 190; Conservative 148; Mismatches 376; Indels 257; Gaps 40;  
QY 16 SKAGPKAGEDEKVAARTRRKVLRAVAATYKMGPAWDOQEYGVSESDGEYAMASS 75  
DB 1023 SKSKVKLEQQVDDLEGSLEQEKVKRMDLERAKRKLQGLD--KLQESIMDLENDKLQLEEK 1081  
QY 76 AESSPGYEWYDEEEKNOLEIE-----RLEEQSLINVDYNCHVDLRLRLRLEGELETK 130  
DB 1082 LKKK-----EPDINQNSKIEDEQALALQJOKLKEN-----QARIEELEELA 1126  
QY 131 VRMARQKMEIPLTELWLEWLBDSMAQDGLDRHVDLFEKAVKDYICPNWLEYG 190  
DB 1127 ETRAKRYEKL-----RSDLSRE----- 1144  
QY 191 QYSGVGGIGQGGLEKVRSVFERA--LSSVGLHMTKGLALWEAYRFPESAIVEAARLEKVH 248  
DB 1145 -----LEEISERLEEAGGATSVQIEMNK-----KREABFQK----- 1175  
QY 249 SILFRROLAIPLYDMEATFAEYEWSESDPIPE--SVIONYNKALQOLEKVK----- 296  
DB 1176 --MRDLLEATIQHEATAALRRKHADSVAELGEQIDNQRVKQLEKESEKLELDDV 1233  
QY 297 --PYEALLQAE-----PRLAEYQVADIFEMKIGDIPARTQILIF--ERALVENCLVPDLMI 348  
DB 1234 TSNMEQIIRKAKANLEKVSRTLEDQAN--EVRVKEEAQRSNLDFTTORAKLOT----- 1284  
QY 349 RYQYLDROLKVKDLVLSVHNRAIRNCPWTVALWSRYL-----LAMRHGVD 395  
DB 1285 --ENGELARQLEKEALISQLTRGKLSYTOQMEDLKRQLEEGKAKNALAHALQASARHDCD 1343

QY 396 ----HQVSVTFEKA-INAGFIQATDYVEIWOAYLDYLRRVDFPKODSSKELEELRAAFTR 451  
Db 1344 LLREQVEETEAKAELQRLVSKANSEVAQW-----RTKYETDAIQRTTELEBAKK 1394  
QY 452 ALEYLQVEEERPNESGDSVIMQWAEIARLNNMQ-----KARE 494  
Db 1395 LAQRL-ODAEAEV-EAVNAKCSLE---KTKHRLQNEIEDLMVDVRSNAAAALDKOR 1449  
QY 495 LWDSIMTRGNKAYNMWLEYINLERAHGDTQHCRCALHRAVQCTSDYPHVCEVLLTMR 554  
Db 1450 NDKILAEWKQYEEQSE---LESSQKARSLSLTKLKNAYESLEH-----LETFR 1502  
QY 555 TEGSLED--WDIAVQ-----KTETRLARVNEQ-RMKAKEAALVQOEEE-KAQQRKRAR 605  
Db 1503 ENKNLQEEISDLTEQLGEGGKNVHELEKVRKQLEVEKLELQSALEAEASLEHEGKIUR 1562  
QY 606 AEKKALKKKKIRGPKRGADEDEKEWG-----DDEEQPSKRRRVENS 650  
Db 1563 AQLEFNQIAKEI---ERKLAERDEEMEQAKRHRVVDLSQTSLSDAETRSRNEVLRVKK 1619  
QY 651 IPAAGETQNVVAAGPAGCAADVPEPPSKQKEKAASLRDMPKVL-----HDSSKDSI 704  
Db 1620 M--EGDLNEMETQLSHANMAA---EAOQVKSQSLKQDQIQIQLDDAVRANDDLKENI 1673  
QY 705 TVFVSNLPSYMOEPTKRLPLFEACGEVVOIRPIFSNRGDFRGYCYVEFEKESALQALE 764  
Db 1674 AIV-----ERRNLLQA--ELEELRAVVEQTERSRLAEQELIETSERVOLLH 1719  
QY 765 MDRKS-VEGRPMFVSPCDKSNPDKVPFVSTSLKHKLFISGLPFSCFCTKEELEICAK 823  
Db 1720 SONTSLINOKKMESDLTQLOVEEAEVQECRABEAKAKAITDAMMAEELKKQDQTS 1779  
QY 824 H-----GTVKDLR-----LVTNRAGPK-----GLAYVEYENESQASOAVMKM 861  
Db 1780 HLERMKKNMEQTIKDLHLDRAEQIALKGGKQKQKLEARVRELEGELEAEQKRAESV 1839  
QY 862 DGMTKENIK 872  
Db 1840 KGMKRSERIK 1850

## RESULT 12

US-09-370-838-183  
; Sequence 183, Application US/09370838  
; Patent No. 6444425

## GENERAL INFORMATION:

; APPLICANT: Reed, Steven G.  
; APPLICANT: Lodes, Michael J.  
; APPLICANT: Mohamath, Roadon  
; APPLICANT: Secrist, Heather  
; TITLE OF INVENTION: COMPOUNDS FOR THERAPY AND DIAGNOSIS OF  
; LUNG CANCER AND METHODS FOR THEIR USE  
; FILE REFERENCE: 210121.475C1  
; CURRENT APPLICATION NUMBER: US/09/370,838  
; CURRENT FILING DATE: 1999-08-09  
; EARLIER APPLICATION NUMBER: US 09/285,323  
; EARLIER FILING DATE: 1999-04-02  
; NUMBER OF SEQ ID NOS: 289  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 183  
; LENGTH: 255  
; TYPE: PRT  
; ORGANISM: Homo sapien

## US-09-370-838-183

Query Match 3.6%; Score 181; DB 4; Length 255;  
Best Local Similarity 23.3%; Pred. No. 2.9e-06;  
Matches 66; Conservative 51; Mismatches 104; Indels 62; Gaps 10;

QY 653 AAGETONVEVAA-----GPGAKCAADVPEPPSKQKEKAASLRD 691  
Db 2 AAGVAAAEVAAETEFKMEESGAPVPSGNGAPGPKG-----EERTONERKKRNIRG 56

QY 692 MPKV-LHDSKSDSIITVFSNLPYSMQEPTDKLRPL-----FEACGEVVOIRPIFSNRGDFR 746  
Db 57 GNRFEPSNPTKRYRAFITNIPF-----DVKQSLKDLVKEKVGVEVYVELLMDABGKSR 111  
QY 747 GYCYVEFEKESALQALE-MDRKSVBGRPMFVSPCVDKSNPDKVFVRYSTSLK-----800  
Db 112 GCVVVEFKMEESKMAAEVINKLSGRPLKV-----KEDPDGEHAR--RAMQKAGRLG 163  
QY 801 HKLFTSGLPFSCFCTKEELEICAKHGTVDKLRVLTNRAGPKGLAYVEYENESQASOAVMK 860  
Db 164 STVFVANLDYKVGWKKLKEVFSMAGVVVRADILEDKGSRGIGVTFEQSIEAQAQISM 223  
QY 861 MDGMTIKENIIKVAISNPPQRPVPEKPETRKAPGGPMLLPQY 903  
Db 224 FNGQLLED-----RPMHVKMDERALPKGDFPPPERH 254

## RESULT 13

US-07-667-276A-7  
; Sequence 7, Application US/07667276A  
; Patent No. 5470971

## GENERAL INFORMATION:

; APPLICANT: Kondo, Keiji  
; APPLICANT: Inouye, Masayori  
; TITLE OF INVENTION: STRESS-INDUCED PROTEINS, GENES CODING  
; THEREFOR, TRANSFORMED CELLS OF ORGANISMS, METHODS AND  
; TITLE OF INVENTION: APPLICATIONS  
; NUMBER OF SEQUENCES: 10  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Weiser & Associates  
; STREET: 230 S. Fifteenth Street, Suite 500  
; CITY: Philadelphia  
; STATE: PA  
; COUNTRY: USA  
; ZIP: 19102

## COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent In Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/07/667,276A  
FILING DATE: 11-MAR-1991

## CLASSIFICATION:

435

## ATTORNEY/AGENT INFORMATION:

NAME: Weiser, Gerard J.  
REGISTRATION NUMBER: 19,763  
REFERENCE/DOCKET NUMBER: 377,5351P

## TELECOMMUNICATION INFORMATION:

TELEPHONE: 215-875-8383

TELEFAX: 215-875-8394

INFORMATION FOR SEQ ID NO: 7:

SEQUENCE CHARACTERISTICS:

LENGTH: 323 amino acids

TYPE: amino acid

TOPOLOGY: unknown

MOLECULE TYPE: protein

US-07-667-276A-7

Query Match 3.6%; Score 178.5; DB 1; Length 323;

Best Local Similarity 26.1%; Pred. No. 6.4e-06;

Matches 61; Conservative 42; Mismatches 88; Indels 43; Gaps 9;

QY 660 VEVAAGPAGKCAADVPEPPSKQKEKAASL-----KRDMPKVLHDSKSDSIITVFSV 709

Db 34 VDVIRIGMTKRGYVDPE-SAEDEKALELTGLKVFNGNEIKLEKPK--GKDSKARTLLAK 90

QY 710 NLPVSMQBPDKRLPLFEACGEVVOIRPIFSNRGDFRGYCYVEFEKESALQALEMDR-K 768

Db 91 NLPYKVTODE--LKEVFEDAAEI-----RLVSKDGSKGLAYIEFTKTEADAEKTFEKG 144

QY 769 SVEGRPMFVSPCVDKSNPDKVFVRYSTSLKHK-----LFTISGLDFSCFCTKEELEICAKH 824

Db 145 EDGR-----SISLYTGERKQNSKTLVLSNYSATEETLQGVFKA 188  
QY 825 GTVKDLRLVTRNAGPKGLAYVEYENESQASQAVNMKGDMGTIKENIKVAISNP 878  
Db 189 TFIK---VPQNGSKSYAFIEFASFEDAKALNSCKNREIEGRAIRLEQGP 239

RESULT 14  
US-07-881-075-3  
; Sequence 3, Application US/07881075  
; Patent No. 544149  
; GENERAL INFORMATION:  
; APPLICANT: KEENE, JACK D.  
; APPLICANT: KING, PETER H.  
; APPLICANT: LEVINE, TODD  
; TITLE OF INVENTION: METHODS AND COMPOSITIONS USEFUL IN THE  
; TITLE OF INVENTION: RECOGNITION, BINDING AND EXPRESSION OF RIBONUCLEIC ACIDS  
; TITLE OF INVENTION: INVOLVED IN CELL GROWTH, NEOPLASIA AND IMMUNOREGULATION  
; NUMBER OF SEQUENCES: 51  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: OBLON, SPIVAK, MCCLELLAND, MAIER & NEUSTADT,  
; STREET: 1755 Jefferson Davis Highway, Fourth Floor  
; CITY: Arlington  
; STATE: Virginia  
; COUNTRY: U.S.A.  
; ZIP: 22202  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/07/881,075  
; FILING DATE: 19920511  
; CLASSIFICATION: 530  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Oblon, No. 544149man F.  
; REGISTRATION NUMBER: 24,618  
; REFERENCE/DOCKET NUMBER: 714-154-0  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (703)521-4500  
; TELEFAX: (703)486-2347  
; TELEX: 248855 OPAT UR  
; INFORMATION FOR SEQ ID NO: 3:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 444 amino acids  
; TYPE: AMINO ACID  
; TOPOLOGY: unknown  
; MOLECULE TYPE: protein  
US-07-881-075-3

Query Match 3.5%; Score 173.5; DB 1; Length 444;  
Best Local Similarity 23.1%; Pred. No. 2.5e-05;  
Matches 81; Conservative 55; Mismatches 122; Indels 93; Gaps 14;

QY 641 PSKRRVENS-IPAAGETQNVAAAGPAGKC--AAVDVEPPSKQKEKAASLRDMPKVLH 697  
Db 34 PATANNVTSQAQTNGGTATTAAAGAGSTTNAAG-----QATANNAASNNNNNNNTN 89  
QY 698 DSSKDSIT-----VFVSNLPYSMQEPDTKLRLPFEACGEVYVQIRPFSN-R 742  
Db 90 NNNNNNATANNNNNEPDKNTLIVNLPQTSQDE--IRSLFVSFGEVSKLIRDKVT 147  
QY 743 GDFRCYCYVEFKEEK-----SALQALEMDRKSVE---GRPMFVSPVDKSNPDKVF 793  
Db 148 GOSLIGYGVXYVKQDAEKAINALNGLRLQNTIKVSIARPS-----189  
QY 794 YSTLEKHKLFISGLPFSCTKEELEICKAHGTVKDLRLVTNR-----AGPKGLAYVE 847  
Db 190 -SESIKANLYVSGLPKNTQSDLESLSFSPYGIITSRLICDNTIDEHAAGLSKGVGFIR 248

QY 848 YENESQASQAVNMKGDMGTIKENIKVAISNPQPKVPKTRKAPGPGMLLPQTYGARG 907  
Db 249 FQDREADRAIKELNGTTPKNS-----TEPTTVFANNPSSNKN-----287  
QY 908 KGRTOISLLRALQPSAAAPQAEENG-----PAAAPAVAAPAPATEAPKMSNA 954  
Db 288 -----SMOPLA-----AYIAPONTGRRAPFANAAAGAAAAAAAIHPNA 328

RESULT 15  
US-08-120-827-3  
; Sequence 3, Application US/08120827  
; Patent No. 5525495  
; GENERAL INFORMATION:  
; APPLICANT: KEENE, JACK D.  
; APPLICANT: KING, PETER H.  
; APPLICANT: LEVINE, TODD  
; TITLE OF INVENTION: METHODS AND COMPOSITIONS USEFUL IN THE  
; TITLE OF INVENTION: RECOGNITION, BINDING AND EXPRESSION OF RIBONUCLEIC ACIDS  
; TITLE OF INVENTION: INVOLVED IN CELL GROWTH, NEOPLASIA AND IMMUNOREGULATION  
; NUMBER OF SEQUENCES: 101  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: OBLON, SPIVAK, MCCLELLAND, MAIER & NEUSTADT,  
; STREET: 1755 Jefferson Davis Highway, Fourth Floor  
; CITY: Arlington  
; STATE: Virginia  
; COUNTRY: U.S.A.  
; ZIP: 22202  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/120,827  
; FILING DATE: 15-SEP-1993  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Oblon, No. 5525495man F.  
; REGISTRATION NUMBER: 24,618  
; REFERENCE/DOCKET NUMBER: 714-158-0 CIP  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (703)413-3000  
; TELEFAX: (703)413-2220  
; TELEX: 248855 OPAT UR  
; INFORMATION FOR SEQ ID NO: 3:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 444 amino acids  
; TYPE: amino acid  
; TOPOLOGY: unknown  
; MOLECULE TYPE: protein  
US-08-120-827-3

Query Match 3.5%; Score 173.5; DB 1; Length 444;  
Best Local Similarity 23.1%; Pred. No. 2.5e-05;  
Matches 81; Conservative 55; Mismatches 122; Indels 93; Gaps 14;

QY 641 PSKRRVENS-IPAAGETQNVAAAGPAGKC--AAVDVEPPSKQKEKAASLRDMPKVLH 697  
Db 34 PATANNVTSQAQTNGGTATTAAAGAGSTTNAAG-----QATANNAASNNNNNNNTN 89  
QY 698 DSSKDSIT-----VFVSNLPYSMQEPDTKLRLPFEACGEVYVQIRPFSN-R 742  
Db 90 NNNNNNATANNNNNEPDKNTLIVNLPQTSQDE--IRSLFVSFGEVSKLIRDKVT 147  
QY 743 GDFRCYCYVEFKEEK-----SALQALEMDRKSVE---GRPMFVSPVDKSNPDKVF 793  
Db 148 GOSLIGYGVXYVKQDAEKAINALNGLRLQNTIKVSIARPS-----189  
QY 794 YSTLEKHKLFISGLPFSCTKEELEICKAHGTVKDLRLVTNR-----AGPKGLAYVE 847  
Db 190 -SESIKANLYVSGLPKNTQSDLESLSFSPYGIITSRLICDNTIDEHAAGLSKGVGFIR 248



Qy	848	YENSQASQAVWKMDGTTIKENIKVAISNPPORKVPEKPETRKAPGGPMLLPQTYGARG	907
Db	249	FDQFEADRAIKELGTTGTPKNS-----TEPIITVKEANNPSSKN-----	287
Qy	908	KGRTQLSLLPRALORPSAAAPQAENG-----PAAAPAAVAAAPAAATEAPKMSNA	954
Db	288	-----SMOPIA-----AYIAQNTGRGRAPFANAAAAGAAAAAAAIHENA	328

Search completed: June 18, 2003, 13:52:47  
Job time : 30 secs

~~THIS PAGE BLANK~~ (USPTO)

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 18, 2003, 13:51:48 ; Search time 55 Seconds  
(without alignments)  
1894.603 Million cell updates/sec

Title: US-09-763-985A-2  
Perfect score: 4994  
Sequence: 1 MATAETSAEPEAEKAGP.....AATEAPKMSNADFAKFLRK 963.

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 41779 seqs, 108206813 residues

Total number of hits satisfying chosen parameters: 41779

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published\_Applications\_AA:\*

- 1: /cgn2\_6/ptodata/2/pubpaa/US08\_NEW\_PUB.pep.\*
- 2: /cgn2\_6/ptodata/2/pubpaa/PCT\_NEW\_PUB.pep.\*
- 3: /cgn2\_6/ptodata/2/pubpaa/US06\_NEW\_PUB.pep.\*
- 4: /cgn2\_6/ptodata/2/pubpaa/US06\_PUBCOMB.pep.\*
- 5: /cgn2\_6/ptodata/2/pubpaa/US07\_NEW\_PUB.pep.\*
- 6: /cgn2\_6/ptodata/2/pubpaa/US07\_PUBCOMB.pep.\*
- 7: /cgn2\_6/ptodata/2/pubpaa/PCTUS\_PUBCOMB.pep.\*
- 8: /cgn2\_6/ptodata/2/pubpaa/US08\_PUBCOMB.pep.\*
- 9: /cgn2\_6/ptodata/2/pubpaa/US09\_NEW\_PUB.pep.\*
- 10: /cgn2\_6/ptodata/2/pubpaa/US09\_PUBCOMB.pep.\*
- 11: /cgn2\_6/ptodata/2/pubpaa/US10\_NEW\_PUB.pep.\*
- 12: /cgn2\_6/ptodata/2/pubpaa/US10\_PUBCOMB.pep.\*
- 13: /cgn2\_6/ptodata/2/pubpaa/US60\_NEW\_PUB.pep.\*
- 14: /cgn2\_6/ptodata/2/pubpaa/US60\_PUBCOMB.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	4989	99.9	1179	10	US-09-821-883-29
2	2918	58.4	578	10	US-09-925-300-1496
3	526	10.5	104	10	US-09-864-761-36026
4	231	4.6	687	9	US-09-969-384-27
5	225	4.5	708	10	US-09-925-301-1012
6	224.5	4.5	714	10	US-09-978-242-3
7	217	4.3	41	10	US-09-864-761-34622
8	208.5	4.2	1938	9	US-10-171-311-164
9	208.5	4.2	1945	9	US-09-927-597-2
10	208.5	4.2	1972	9	US-10-171-311-162
11	208.5	4.2	1979	9	US-09-927-597-4
12	206.5	4.1	494	10	US-09-833-790-234
13	199.5	4.0	852	10	US-09-752-639-153
14	199.5	4.0	852	10	US-09-984-198-153
15	199	4.0	687	9	US-09-969-384-16
16	194	3.9	446	10	US-09-925-300-1554
17	188.5	3.8	1203	9	US-10-097-340-43
18	186.5	3.7	482	9	US-10-153-668-320
19	184.5	3.7	486	9	US-10-153-668-432

20	182.5	3.7	2871	9	US-10-146-473-41	Sequence 41, Appl
21	181.5	3.6	2383	9	US-10-082-830-260	Sequence 260, App
22	181	3.6	255	9	US-09-854-133-183	Sequence 183, App
23	181	3.6	255	10	US-09-738-973-183	Sequence 183, App
24	181	3.6	2125	10	US-09-919-172-29	Sequence 29, Appl
25	172	3.4	2462	9	US-09-819-104A-5	Sequence 5, Appl
26	171	3.4	633	9	US-09-374-046A-128	Sequence 128, App
27	171	3.4	633	10	US-09-821-687-10	Sequence 10, Appl
28	170	3.4	380	9	US-10-025-367-27	Sequence 27, Appl
29	170	3.4	484	9	US-10-153-668-388	Sequence 388, App
30	170	3.4	488	9	US-10-153-668-392	Sequence 392, App
31	170	3.4	509	9	US-10-153-668-272	Sequence 272, App
32	167.5	3.4	397	9	US-10-106-698-4611	Sequence 4611, Ap
33	166.5	3.3	691	9	US-10-028-072-16	Sequence 16, Appl
34	166.5	3.3	691	9	US-10-121-049-16	Sequence 16, Appl
35	166.5	3.3	691	9	US-10-123-904-16	Sequence 16, Appl
36	166.5	3.3	691	9	US-10-140-470-16	Sequence 16, Appl
37	166.5	3.3	691	9	US-10-175-746-16	Sequence 16, Appl
38	166.5	3.3	691	9	US-10-176-918-16	Sequence 16, Appl
39	166.5	3.3	691	9	US-10-176-921-16	Sequence 16, Appl
40	166.5	3.3	691	9	US-10-137-865-16	Sequence 16, Appl
41	166.5	3.3	691	9	US-10-140-474-16	Sequence 16, Appl
42	166.5	3.3	691	9	US-10-142-431-16	Sequence 16, Appl
43	166.5	3.3	691	9	US-10-143-114-16	Sequence 16, Appl
44	166.5	3.3	691	9	US-10-140-002-16	Sequence 16, Appl
45	166.5	3.3	691	9	US-10-142-419-16	Sequence 16, Appl

#### ALIGNMENTS

RESULT 1  
US-09-821-883-29  
; Sequence 29, Application US/09821883  
; Patent No. US20020061310A1  
; GENERAL INFORMATION:  
; APPLICANT: Laus, Reiner  
; APPLICANT: Graddis, Thomas  
; APPLICANT: Vidovic, Damir  
; TITLE OF INVENTION: Compositions and Methods for Dendritic  
; TITLE OF INVENTION: Cell-Based Immunotherapy  
; FILE REFERENCE: 7636-0022.30  
; CURRENT APPLICATION NUMBER: US/09/821,883  
; CURRENT FILING DATE: 2001-03-30  
; PRIOR APPLICATION NUMBER: US 60/193,504  
; PRIOR FILING DATE: 2000-03-30  
; NUMBER OF SEQ ID NOS: 30  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 29  
; LENGTH: 1179  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: SART-3-IC  
US-09-821-883-29

Query Match	99.9%	Score 4989;	DB 10;	Length 1179;
Best Local Similarity	100.0%;	Pred. No. 2.7e-309;	Mismatches 0;	Indels 0;
Matches 962;	Conservative 0;			Gaps 0;
QY	1	MATAETSAEPEAEKAGPKADGEDEVKAAARTRRKVLRAVAATAATYKTGMPADQDEE 60		
Db	1	MATAETSAEPEAEKAGPKADGEDEVKAAARTRRKVLRAVAATAATYKTGMPADQDEE 60		
QY	61	GVSESDGDEYAMASSAESPGYEYWEYDEEEKNOLETERLEEQLSINVDYNCHVDLIR 120		
Db	61	GVSESDGDEYAMASSAESPGYEYWEYDEEEKNOLETERLEEQLSINVDYNCHVDLIR 120		
QY	121	LRLRLEGETTKVRMAQKMSIEIPLTFELWLEWHDIEISMAQGLDREHYVDLFEKAVKDY 180		
Db	121	LRLRLEGETTKVRMAQKMSIEIPLTFELWLEWHDIEISMAQGLDREHYVDLFEKAVKDY 180		
QY	181	ICPNLWLEYGOYSVGGIGQKGLGKVRSVFERALSSVGLHMTKGLALWEAYREFESAIVE 240		

Db 181 ICNINILEYGOYVGGIGQGGLEKRVSVFERALSSVGLHMTKGLALWEAYREFESAIVE 240  
QY 241 AARLEKVVHSFRRLQALPIYDMETAEAEVEESEDPIPESVIONYKALQOLEKYPYEE 300  
Db 241 AARLEKVVHSFRRLQALPIYDMETAEAEVEESEDPIPESVIONYKALQOLEKYPYEE 300  
QY 301 ALLOAEAPRLAEOAYIDFEMKIGDPARIQIFERALVENCNCLVDPDLWIRYSQYLDRLQKV 360  
Db 301 ALLOAEAPRLAEOAYIDFEMKIGDPARIQIFERALVENCNCLVDPDLWIRYSQYLDRLQKV 360  
QY 361 KDILVSVHNRAINCPTWALMSRYLLAMERHGVHDQVTSVTEKALNAGFQATDYVEI 420  
Db 361 KDILVSVHNRAINCPTWALMSRYLLAMERHGVHDQVTSVTEKALNAGFQATDYVEI 420  
QY 421 WQAYLDYLRVRVDFKODSSKELEELRAAFTRALEYLKQVEEVEEVEEVEEVEEVEEVEE 480  
Db 421 WQAYLDYLRVRVDFKODSSKELEELRAAFTRALEYLKQVEEVEEVEEVEEVEEVEEVEE 480  
QY 481 IEARLCNNMOKARELWDSIMTRGNAKYANNMWEYNNLERAHGDTQHCRCALHRAVQCTSD 540  
Db 481 IEARLCNNMOKARELWDSIMTRGNAKYANNMWEYNNLERAHGDTQHCRCALHRAVQCTSD 540  
QY 541 YPEHVCEVLLTMERTEGSLEDWDIAVQKTETRLARVNEQRMKAEEAALVQOEEKAEQ 600  
Db 541 YPEHVCEVLLTMERTEGSLEDWDIAVQKTETRLARVNEQRMKAEEAALVQOEEKAEQ 600  
QY 601 RKRAAEKALKKKKIRGPEKRGADDEDEKQWDDDEEOPSKRRRVENSIPAAGETQNV 660  
Db 601 RKRAAEKALKKKKIRGPEKRGADDEDEKQWDDDEEOPSKRRRVENSIPAAGETQNV 660  
QY 661 EVAAGPAGCAAVDVPEPPSKQKEKAASLRDMPKVLHDSKSDSITVFSNLPYSMOEPTD 720  
Db 661 EVAAGPAGCAAVDVPEPPSKQKEKAASLRDMPKVLHDSKSDSITVFSNLPYSMOEPTD 720  
QY 721 KLRLPEACEGVVQIRPIFSNRGDFRGYCYVEFEKESALQALEMDRKSVEGRPMFVSPC 780  
Db 721 KLRLPEACEGVVQIRPIFSNRGDFRGYCYVEFEKESALQALEMDRKSVEGRPMFVSPC 780  
QY 781 VDKSKNPDKFVFRYSTSLEKHKLFISGLPFSCCTKELEICKAHGTVKDLRLVTNRAGKP 840  
Db 781 VDKSKNPDKFVFRYSTSLEKHKLFISGLPFSCCTKELEICKAHGTVKDLRLVTNRAGKP 840  
QY 841 KGLAYVEYENESQASQAVMKMDGWTIKENIIKVAISNPPQKVPKPEPTRKAPGGPMLLP 900  
Db 841 KGLAYVEYENESQASQAVMKMDGWTIKENIIKVAISNPPQKVPKPEPTRKAPGGPMLLP 900  
QY 901 QTYGARGKGTQSLPRALQPSAAPOAENGPAAPAAVAAAPAAAPAAAPAAAPAAAPAA 960  
Db 901 QTYGARGKGTQSLPRALQPSAAPOAENGPAAPAAVAAAPAAAPAAAPAAAPAAAPAA 960  
QY 961 LR 962  
Db 961 LR 962

RESULT 2

US-09-925-300-1496  
; Sequence 1496, Application US/09925300  
; Patent No. US20020151681A1  
; GENERAL INFORMATION:  
; APPLICANT: Craig Rosen,  
; APPLICANT: Steve Ruben  
; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies  
; FILE REFERENCE: PA101  
; CURRENT APPLICATION NUMBER: US/09/925,300  
; CURRENT FILING DATE: 2001-08-10  
; PRIOR APPLICATION NUMBER: PCT/US00/05988  
; PRIOR FILING DATE: 2000-03-08  
; PRIOR APPLICATION NUMBER: 60/124,270  
; PRIOR FILING DATE: 1999-03-12  
; NUMBER OF SEQ ID NOS: 1890  
; SOFTWARE: Patentin Ver. 2.0

; SEQ ID NO 1496  
; LENGTH: 578  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-925-300-1496

Query Match 58.4%; Score 2918; DB 10; Length 578;  
Best Local Similarity 99.1%; Pred. No. 7.3e-178;  
Matches 564; Conservative: 0; Mismatches 5; Indels 0; Gaps 0;

QY 395 DHOVIVSVTEKALNAGFQATDYVEITWQAYLDYLRVRVDFKQDSSKELEELRAAFTRALE 454  
Db 10 DESPTPTTEFKALNAGFQATDYVEITWQAYLDYLRVRVDFKQDSSKELEELRAAFTRALE 69  
QY 455 YLKQVEEVEEVEEVEEVEEVEEVEEVEEVEEVEEVEEVEEVEEVEEVEEVEEVEE 514  
Db 70 YLKQVEEVEEVEEVEEVEEVEEVEEVEEVEEVEEVEEVEEVEEVEEVEEVEEVE 129  
QY 515 YNLERAHGDTHCRKALHRAVQCTSDYPEHVCEVLLTMERTEGSLEDWDIAVQKTETRLA 574  
Db 130 YNLERAHGDTHCRKALHRAVQCTSDYPEHVCEVLLTMERTEGSLEDWDIAVQKTETRLA 189  
QY 575 RVNEQRMKAEEAALVQOEEKAEQRRARAEEKALKKKKIRGPEKRGADDEDEKQW 634  
Db 190 RVNEQRMKAEEAALVQOEEKAEQRRARAEEKALKKKKIRGPEKRGADDEDEKQW 249  
QY 635 DDEEOPSKRRRVENSIPAAGETQNVVEAAGPAGCAAVDVPEPPSKQKEKAASLRDMPK 694  
Db 250 DDEEOPSKRRRVENSIPAAGETQNVVEAAGPAGCAAVDVPEPPSKQKEKAASLRDMPK 309  
QY 695 VLHDSKSDSITVFSNLPYSMOEPTDKLRLPEACEGVVQIRPIFSNRGDFRGYCYVEFK 754  
Db 310 VLHDSKSDSITVFSNLPYSMOEPTDKLRLPEACEGVVQIRPIFSNRGDFRGYCYVEFK 369  
QY 755 EKESALQALEMDRKSVEGRPMFVSPCVDKSKNPDKFVFRYSTSLEKHKLFISGLPFSCCTK 814  
Db 370 EKESALQALEMDRKSVEGRPMFVSPCVDKSKNPDKFVFRYSTSLEKHKLFISGLPFSCCTK 429  
QY 815 EELEICKAHGTVKDLRLVTNRAGPKGLAYVEYENESQASQAVMKMDGWTIKENIIKVA 874  
Db 430 EELEICKAHGTVKDLRLVTNRAGPKGLAYVEYENESQASQAVMKMDGWTIKENIIKVA 489  
QY 875 ISNPPQKVPKPEPTRKAPGGPMLLPQTYGARGKGTQSLPRALQPSAAPOAENGPA 934  
Db 490 ISNPPQKVPKPEPTRKAPGGPMLLPQTYGARGKGTQSLPRALQPSAAPOAENGPA 549  
QY 935 AAAPAAVAAAPAAAPAAAPAAAPAAAPAAAPAAAPAAAPAAAPAAAPAAAPAA 963  
Db 550 AAAPAAVAAAPAAAPAAAPAAAPAAAPAAAPAAAPAAAPAAAPAAAPAAAPAA 578

RESULT 3

US-09-864-761-36026  
; Sequence 36026, Application US/09864761  
; Patent No. US20020048763A1  
; GENERAL INFORMATION:  
; APPLICANT: Penn, Sharron G.  
; APPLICANT: Rank, David R.  
; APPLICANT: Hanzel, David K.  
; APPLICANT: Chen, Wensheng  
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FO  
; FILE REFERENCE: Aeo mica-x-1  
; CURRENT APPLICATION NUMBER: US/09/864,761  
; CURRENT FILING DATE: 2001-05-23  
; PRIOR APPLICATION NUMBER: US 60/180,312  
; PRIOR FILING DATE: 2000-02-04  
; PRIOR APPLICATION NUMBER: US 60/207,456  
; PRIOR FILING DATE: 2000-05-26  
; PRIOR APPLICATION NUMBER: US 09/632,366  
; PRIOR FILING DATE: 2000-08-03  
; PRIOR APPLICATION NUMBER: GB 24263.6  
; PRIOR FILING DATE: 2000-10-04



```
; Sequence 1012, Application US/09925301
; Patent No. US20020052308A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
; FILE REFERENCE: PA106
; CURRENT APPLICATION NUMBER: US/09/925,301
; CURRENT FILING DATE: 2001-08-10
; PRIOR APPLICATION NUMBER: PCT/US00/05882
; PRIOR FILING DATE: 2000-03-08
; PRIOR APPLICATION NUMBER: 60/124,270
; PRIOR FILING DATE: 1999-03-12
; NUMBER OF SEQ ID NOS: 1694
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1012
; LENGTH: 708
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (18)
; OTHER INFORMATION: xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (153)
; OTHER INFORMATION: xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (229)
; OTHER INFORMATION: xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (433)
; OTHER INFORMATION: xaa equals any of the naturally occurring L-amino acids
; US-09-925-301-1012

Query Match
Best Local Similarity 4.5%; Score 225; DB 10; Length 708;
Matches 145; Conservative 125; Mismatches 279; Indels 188; Gaps 34;

QY 14 AFSKAGKADGDEVEDEKAAARTKV-----LSRAVAAATYKMGPAWDQOEEGVS 63
Db 19 AADMAASTAGQRIKPKYAKVKNKAPAEVQITAEQLLREAKERELELLPP--PPQKID 76
QY 64 ESDGDEYAMASAESPGYEYDEEKNQLEI-----ERLEE-QLSINVYD 112
Db 77 EELNDYKLRKRK-----TFEDNIRKNTVISNWKYAAQWESLKEIQARSIVER 127
QY 113 NCHVDL-----IRLLRLEGELTKVRMAR--QKXSEIFPLTEELWLEWHD-----IS 158
Db 128 ALDVIDYRNTLWLYKAEMEMKRNQVHAHNIWDRAITTLPRVYNQFYKITYMEMLGNVA 187
QY 159 MAQDGLDRHVDLFEKAVKDYICPNWLEYQYSGVGIGQKGLKVRVSPERALS SVG 218
Db 188 GARQVFERMWEQPEQAWHSYI--NFELRYKE-----VDRARTIYER--XVL 231
QY 219 LHMTGLALWEAYREFESAIVAARLEKVKHSFLRQLAIPLYDME-----ATAEYEE 271
Db 232 VH--PDVKNWIKYAFEEKHAYFAHARKYI-----ERAVEFFGDEHMDHLELVAFAPKEE 284
QY 272 WSEDPTPEVIONKALQLEKYPYEEALLQAEAPRLAEOAYIDFEMKIGDPARQL 331
Db 285 NQKEPERVVIKY--ALDRISKQDAQE-----LFKNYTFEKKFGDRRGIED 330
QY 332 I-----FERALVENCIVLPDLWTRYQYLDRLQKVDLVLVSVINRAIRNC----- 376
Db 331 IIVSKRRFOYEEVKANPHNYDAWFDYLRVLSDEAE--EAVREYVERAIANVPPIQEKRH 389
QY 377 WT--VALMSRYLLAMERHGVHDQVLSVTFEKNALNAGFTQADTVELWQAYLDLRRVDF 434
Db 390 WKRYIYLWNYALYEELEAKDPTRQVYQASLELPHKFTFKXKMWILYAFQEIQR--- 446
QY 435 KQDSSKELEELRAAFTRAL-----EYLUKQVE-----ERFNESGDPSCV 473
Db 447 -----KNLSLARRALGTSIGKCPKNKLFKVYIELEQLREFDRCKRLEKFELEFGPNC 501
```

```
QY 474 IMQNWARIEARLNCNNQKARELWDSIMTRGNKAYAN--MWLEYNNLERAHGDTHCKRKAH 532
Db 502 SWIKFAELET--ILGDIDRARIYELAISQRLDMPVLWKSVIDFEIEQEEETRTNLYR 560
QY 533 RAVQCTSDYPHEVCEVLLTMR-----TEGSLEWDIDIAVQKTETRLARVNQRMK---- 582
Db 561 RLLQRT---QHV-KVWISFAQFELSSGKESL-----TKRQIYEANKTMRN 604
QY 583 AAEKEAALV-----QOEEKAEQKRARAER---KALKKKKKIRGPKRGADEDEDEKEM 633
Db 605 CEEKERMLLESWRSFEEFCTASDKERVDKLMPEKVKRRKVQ-----TDDGSDAGW 658
QY 634 -----GDDEEQPS 642
Db 659 BEYFDYIRPEDAANQPN 675

RESULT 6
US-09-978-242-3
; Sequence 3, Application US/09978242
; Patent No. US20020098566A1
; GENERAL INFORMATION:
; APPLICANT: Bandman, Olga
; Yue, Henry
; Corley, Neil C.
; Shah, Purvi
; TITLE OF INVENTION: HUMAN NUCLEOLIN-LIKE PROTEIN
; NUMBER OF SEQUENCES: 3
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Incyte Pharmaceuticals, Inc.,
; STREET: 3174 Porter Drive
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/978,242
; FILING DATE: 15-Oct-2001
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 09/241,333
; FILING DATE: <Unknown>
; APPLICATION NUMBER: 08/990,114
; FILING DATE: <Unknown>
; ATTORNEY/AGENT INFORMATION:
; NAME: Billings, Lucy J.
; REGISTRATION NUMBER: 36,749
; REFERENCE/DOCKET NUMBER: PF-0451 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 650-855-0555
; TELEFAX: 650-845-4166
; TELEX: <Unknown>
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 714 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; LIBRARY: GenBank
; CLONE: 128842
; SEQUENCE DESCRIPTION: SEQ ID NO: 3:
US-09-978-242-3
```

```
Query Match
Best Local Similarity 4.5%; Score 224.5; DB 10; Length 714;
Matches 94; Conservative 67; Mismatches 134; Indels 105; Gaps 16;
```

Qy	583	AAEKAAALVQEEEAQAORRKARAB-----KKALKKKKKIRGPEKKGAEDDE	630
Db	187	ASEDEDEEEEEDEEDEDSEEEAEMBITPAKKAKAPAKVPVKAKNVAEDDDDD	246
Qy	631	KWGGDDEEQ-----PSKRRR---VENSIPAAGTQNV---	661
Db	247	EDEDEDEDEDEDEDEDEDEDEEPVKPAGARKKEMTKQKEVPEA-KKQVEGEST	305
Qy	662	-----VAAGPAGK--CAAVDVPPSPSK-	686
Db	306	TPPNLFIGNLNPNSVAELKAISPEPAKNDLAVVDVTCTNRKFGYVDFESAEDLEKAL	365
Qy	687	SL-----KRMPKVLHDSK--DSITVFVSNLPLYSMQBPDTKLRLPLFACGEVVQ	734
Db	366	EUTGLTVFGNEIKLEPKPG-RDSKKVRAARTLLAKNISFNITEDE--LKVEFF-DALE	419
Qy	735	IRPIFSNGDRFGYCYVEFFEKSALQALEMDR-KSVEGRPMFVSPCVDKSNKPDPKFVR	793
Db	420	IR-LYSQDGKSGIAYIEFKESADAKNLEBKQAEIDGRSVLSYYTGKGQQRGTGN	478
Qy	794	YSTLESKHKLFTISGLFPFSCTEEELEEICAHGTVKDLRLVTNRAGPKGLIAYEYENESQ	853
Db	479	STWSGESKTVLVLSNLSYSATEETLQEVFEKATFIK---VPQNQQGKSGVAFTEFASFED	535
Qy	854	ASQAVMKMDGWTIKENIITKVAINPPORKYPEKETRAP	893
Db	536	AREALNSCNMBIEGRTIRLEIQGP-----RGSPNARSOP	570

## RESULT 7

```

US-09-864-761-34622
; Sequence 34622, Application US/09864761
; Patent NO. US20020048763A1
GENERAL INFORMATION:
; APPLICANT: Penn, Sharron G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
; FILE REFERENCE: Aemica-X-1
; CURRENT APPLICATION NUMBER: US/09/864,761
; CURRENT FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: GB 24263.6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30

```

```

; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; PRIOR FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 49117
; SOFTWARE: Annonax Sequence Listing Engine vers. 1.1
; SEQ ID NO 34622
; LENGTH: 41
; TYPE: PR1
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AC008119.6
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 1.6
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 1.2
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 1.9
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 1.4
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 1.2
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 1.8
; OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 1.5
; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 1.6
; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 1.7
; OTHER INFORMATION: EXPRESSED IN HBL100, SIGNAL = 1.9
; OTHER INFORMATION: SWISSPROT HIT: Q32866, EVALU6 6.10e+00
; OTHER INFORMATION: EST_HUMAN HIT: AU143744.1, EVALU6 1.00e-17
; US-09-864-761-34622

```

Query Match 4.3%; Score 217; DB 10; Length 41;  
Best Local Similarity 100.0%; Pred. No. 2.2e-07;  
Matches 41; Conservative 0; Mismatches 0 Indels

QY	262	MEATFAEYE	EWESED	PIPESVIQ	NYNKALQ	LEKYKPY	EAL	302
Db	1	MEATFAEYE <td>EWESED <td>PIPESVIQ <td>NYNKALQ <td>LEKYKPY <td>EAL</td> <td>41</td> </td></td></td></td>	EWESED <td>PIPESVIQ <td>NYNKALQ <td>LEKYKPY <td>EAL</td> <td>41</td> </td></td></td>	PIPESVIQ <td>NYNKALQ <td>LEKYKPY <td>EAL</td> <td>41</td> </td></td>	NYNKALQ <td>LEKYKPY <td>EAL</td> <td>41</td> </td>	LEKYKPY <td>EAL</td> <td>41</td>	EAL	41

## RESULTS

```

RESUL1 6
; Sequence 164, Application US/10171311
; Publication No. US20030087270A1
; GENERAL INFORMATION:
; APPLICANT: Schlegel, Robert
; APPLICANT: Chen, Yan
; APPLICANT: Zhao, Xumei
; APPLICANT: Monahan, John
; APPLICANT: Kamatkar, Shubhangi
; APPLICANT: Glatt, Karen
; APPLICANT: Gannavarapu, Manjula
; APPLICANT: Hoersh, Sebastian
; TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND METHODS FOR
; TITLE OF INVENTION: IDENTIFICATION, ASSESSMENT, PREVENTION, AND THERAPY
; TITLE OF INVENTION: OF CERVICAL CANCER
; FILE REFERENCE: MRI-035
; CURRENT APPLICATION NUMBER: US/10/171,311
; CURRENT FILING DATE: 2002-06-12
; PRIOR APPLICATION NUMBER: US 60/298,159
; PRIOR FILING DATE: 2001-06-13
; PRIOR APPLICATION NUMBER: US 60/298,155
; PRIOR FILING DATE: 2001-06-13
; PRIOR APPLICATION NUMBER: US 60/335,936
; PRIOR FILING DATE: 2001-11-14
; NUMBER OF SEQ ID NOS: 238
; SOFTWARE: FastSeq for Windows version 4.0
; SEQ ID NO 164
; LENGTH: 1938
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-171-311-164

```

Query Match	4.2%	Score 208.5;	DB 9;	Length 1938;
Best Local Similarity	19.6%	Pred. No. 0.0001;		
Matches 213:	Conservative 170;	Mismatches 399;	Indels 305;	Gaps 44;





Db 1485 AREKTRALSARALEEAEKEELERTNKLKAEEMEDLVSKDDVGNVHELEKSKRAL 1544  
Qy 550 LT-MERTEGSLDWDIAVOKTETRLAR--VNEORMKA-----AKEAALVOQ 593  
Db 1545 ETQMEEMKTQLEEDLEQATEADAKLRLEVNQALKGOFERDQARDEQNEEKRRQORQ 1604  
Qy 594 -----EEKABQQRKRAAEK-----ALKKKKIRGPEKRG 624  
Db 1605 LHEYTELEDERKQALAAAKKLEGLDKDLELQADSIAIKGREAIKQLRKLQAKMDF 1664  
Qy 625 ADEDEKEWGDDE-----EEQPSKRRRVENSTIPAAGETQNVVAAGPAGKCAAVDVEPPS 679  
Db 1665 QRELEDRASRDEIFATAKENEKKAKEADLMQLE-----DLAAERARKQA-DLEKEE 1719  
Qy 680 KOKEKAASLRDMPKVLHDSKDSITVFNLSNLPYSMOEPDKLRPLFEACGEVQIRPIF 739  
Db 1720 LAELASSLSG--RNAQD-EKRRLAARIAQLEEEEEEQGNMEAMSDRVKATQQAOL 1776  
Qy 740 SNRGDFRGYCYVEFEKESALQALEMDRKSVEGRPMFVSPCVDKSKNPDVFKVFRYSTSLE 799  
Db 1777 SN-----ELATERSTAQKNESARQOLERO-----1800  
Qy 800 KHLFISGLPSCCTKEELEECAGHTVVDKDLRLVTNRAGKPKGLAYVE-----NESQA 854  
Db 1801 -----NKLRSKLHEMEGAVKSKFKSTIAALEAK-IAQLEEQVEQEAEREKQA 1846  
Qy 855 SOAVMKMDGWTIKENIKVAISNPQRKVPETPKAPGGPMLLPQTYGARGKGTOLS 914  
Db 1847 ATKSLKQDKKILKEILLQV-----EDERKMAEQYKEQAB-----KGNARVK 1887  
Qy 915 LLPRALQ 921  
Db 1888 QLKROLE 1894

## RESULT 10

US-10-171-311-162  
; Sequence 162, Application US/10171311  
; Publication No. US20030087270A1  
; GENERAL INFORMATION:  
; APPLICANT: Schlegel, Robert  
; APPLICANT: Chen, Yan  
; APPLICANT: Zhao, Xumei  
; APPLICANT: Monahan, John  
; APPLICANT: Kamatkar, Shubhangi  
; APPLICANT: Glatt, Karen  
; APPLICANT: Gannavarapu, Manjula  
; APPLICANT: Hoerish, Sebastian  
; TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND METHODS FOR  
; TITLE OF INVENTION: IDENTIFICATION, ASSESSMENT, PREVENTION, AND THERAPY  
; TITLE OF INVENTION: OF CERVICAL CANCER  
; FILE REFERENCE: MRI-035  
; CURRENT APPLICATION NUMBER: US/10/171,311  
; CURRENT FILING DATE: 2002-06-12  
; PRIOR APPLICATION NUMBER: US 60/298,159  
; PRIOR FILING DATE: 2001-06-13  
; PRIOR APPLICATION NUMBER: US 60/298,155  
; PRIOR FILING DATE: 2001-06-13  
; PRIOR APPLICATION NUMBER: US 60/335,936  
; PRIOR FILING DATE: 2001-11-14  
; NUMBER OF SEQ ID NOS: 238  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 162  
; LENGTH: 1972  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-10-171-311-162

Query Match 4.2%; Score 208.5; DB 9; Length 1972;  
Best Local Similarity 19.6%; Pred. No. 0.0001;  
Matches 213; Conservative 170; Mismatches 399; Indels 305; Gaps 44;

Qy 6 ETSASEPDAESK--AGPKADGEE--DEVKAAATRRKVLRSRAVAATYKTMGP---AWDQO 58  
Db 935 EDRGQLQAEKKMAQQMLDLEQLEEEAEARQKIQ-LEKVTAAEKIKKLEDEILVMDQ 993  
Qy 59 -----BEGVSE-----SOGDEYA-----MASSAESPGYEWYDEEBEKNQLE 97  
Db 994 NNKLSKERKLLERISDLTTNLAEEERAKNLTKLKNKHESMISELYRL-KKEEKROE 1052  
Qy 98 IERLEEQLSINVDYNCHVDLRLRLRGELTKVMAQKMEIPPLFEELWLEWHDEI 157  
Db 1053 LEKLRKLEGASDPH-----EQIADQQAELAKMLAKKEEL-----QAALRLDDEI 1103  
Qy 158 SMAQDGLDR-----EHVYDLFEKAVKDYICPNWILEYQYSGVGIGKGLKVRSVFER 212  
Db 1104 AQKNNAKKIRELEGHISDLQEDLDSARAARN---KAEKOKRDLGEE--LEALTELED 1157  
Qy 213 ALSSVGLHMTKGLALWEAYREFESAIVEAARLEKVHSLFRRLQALPLVDMEATFAYEEM 272  
Db 1158 TLDSTATQOE-----LRAKREQEVTVLKALDEETRS-----HEAQVQEMROK 1200  
Qy 273 SEDPIESVIQ-----NYNKALQOLEKYKPYEEALLQAEAPRLAEYQAYIDPEMK- 322  
Db 1201 HAQAVEELTEQEQFKRAKANLDKNKQTLK-----ENADLAGELRVLGAKQVEVHKKK 1256  
Qy 323 -----IGDPARIQIIFERALVEN-----CLVPDLWIRYSQ 352  
Db 1257 LEAQVOELQSKSDGERARAEALNDKVHKLQNEVESVTGMLNAECAIKLAKDVASLSQ 1316  
Qy 353 YLDROLKVKD-----LVLSVHNRAIRNCPWTVALWSRYLLAMERHGVHDHV-----ISVT 402  
Db 1317 LQDTQELLEQEEETROKLVNSTKLQLEE-----ERNSIQDQLDEEMEAKON 1361  
Qy 403 FEKALNAGFIQATDYVEIWOAYLDYLRRRVDFKODSSKELELRAAF--TRALEYLK-QEV 460  
Db 1362 LERHISTLNQLSDSKKKLQDFASTVEALEBEGKKRFQKEIENLTQOYEKAAYDKLEKT 1421  
Qy 461 EEFNESGDPSCVIMONWARIARLNCNMKARELWDSIMT-----RGNAK 506  
Db 1422 KNRLQOELDLVLDLN---QRLVSNLEKKORKKFDOLLAEBEKNISSYADERDRAEAE 1477  
Qy 507 YANMWLEYNLERAHGDTQHCRAKHRA-----VQCTSDYPEHVCEV-----L 549  
Db 1478 AREKETKALSARALEEAELEAKEELERTNKLKAEEMEDLVSKDDVGNVHELEKSKRAL 1537  
Qy 550 LT-MERTEGSLDWDIAVOKTETRLAR--VNEORMKA-----AKEAALVOQ 593  
Db 1538 ETQMEEMKTQLEEDLEQATEADAKLRLEVNQALKGOFERDQARDEQNEEKRRQORQ 1597  
Qy 594 -----EEKAEQQRKRAAEK-----ALKKKKIRGPEKRG 624  
Db 1598 LHEYTELEDERKQALAAAKKLEGLDKDLELQADSIAIKGREAIKQLRKLQAKMDF 1657  
Qy 625 ADEDEKEWGDDE-----EEQPSKRRRVENSTIPAAGETQNVVAAGPAGKCAAVDVEPPS 679  
Db 1658 QRELEDRASRDEIFATAKENEKKAKEADLMQLE-----DLAAERARKQA-DLEKEE 1712  
Qy 680 KOKEKAASLRDMPKVLHDSKDSITVFNLSNLPYSMOEPDKLRPLFEACGEVQIRPIF 739  
Db 1713 LAELASSLSG--RNAQD-EKRRLAARIAQLEEEEEEQGNMEAMSDRVKATQQAOL 1769  
Qy 740 SNRGDFRGYCYVEFEKESALQALEMDRKSVEGRPMFVSPCVDKSKNPDVFKVFRYSTSLE 799  
Db 1770 SN-----ELATERSTAQKNESARQOLERO-----1793  
Qy 800 KHLFISGLPSCCTKEELEECAGHTVVDKDLRLVTNRAGKPKGLAYVE-----NESQA 854  
Db 1794 -----NKLRSKLHEMEGAVKSKFKSTIAALEAK-IAQLEEQVEQEAEREKQA 1839  
Qy 855 SOAVMKMDGWTIKENIKVAISNPQRKVPETPKAPGGPMLLPQTYGARGKGTOLS 914  
Db 1840 ATKSLKQDKKILKEILLQV-----EDERKMAEQYKEQAB-----KGNARVK 1880  
Qy 915 LLPRALQ 921







```

542 Qy -----PEHV-----CEVLLTWTERTSGSLSDWDIAVOKT 569
      :||
546 Db -QHVWYISFAQFELSSRRKKFAKQIYEKLTKPCE--TVKKRRDLCCWN----- 595
      :||
570 Qy ETRLRVNEORMKAAEKEAALVQOEEKABQRKAR----- 605
      |||
596 Db ---LGEVLEEFGTASDKERVVDLMPKVKRRKVQPDGSDAGWEYFYIFEDAANQ 652
      :||
606 Qy -----AEKALKKKKKIRGPEKGADED--DEKE 632
      |||
653 Db PNKLAMAKLWKQOQKEGDABHHPDDEYDESE 686
      |||

```

Search completed: June 18, 2003, 14:01:09  
Job time : 60 secs

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 18, 2003, 13:48:53 ; Search time 26 Seconds  
(without alignments)  
3560.671 Million cell updates/sec

Title: US-09-763-985A-2

Perfect score: 4994

Sequence: 1 MATAAETSASEPEAESKACP.....AATEAPKMSNADFAKLFLRK 963

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

PIR\_73:\*

1: pir1:\*

2: pir2:\*

3: pir3:\*

4: pir4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	864	17.3	836	2	T01563
2	823	16.5	768	2	T09890
3	506.5	10.1	1014	2	T39743
4	305	6.1	651	2	S18874
5	293.5	5.9	705	2	S32644
6	276.5	5.5	761	2	T31808
7	272	5.4	414	2	A39205
8	268.5	5.4	702	2	T13427
9	262	5.2	694	1	DNCNHL
10	245.5	4.9	707	2	A35804
11	244.5	4.9	707	1	DNMS
12	243	4.9	557	2	A96527
13	239.5	4.8	960	2	T17297
14	226.5	4.5	521	2	A40252
15	224.5	4.5	713	2	A27441
16	224	4.5	712	2	JH0148
17	222.5	4.5	599	2	T45742
18	217.5	4.4	1938	2	JC5421
19	217.5	4.4	1972	2	JC5420
20	216.5	4.3	717	2	S50852
21	216	4.3	492	2	S22646
22	216	4.3	1130	2	T34081
23	215	4.3	733	2	T11668
24	214.5	4.3	611	2	T06458
25	212.5	4.3	524	2	I55595
26	211	4.2	635	2	T09648
27	210	4.2	483	2	A33130
28	209.5	4.2	500	2	S55785
29	209	4.2	978	2	A70387

30 207 4.1 521 2 T01563  
31 207 4.1 1959 1 A33977  
32 206 4.1 315 2 S22548  
33 205.5 4.1 597 2 T47673  
34 205.5 4.1 1972 1 A41604  
35 204.5 4.1 289 2 S50765  
36 203.5 4.1 604 2 T37994  
37 202.5 4.1 1937 2 I38055  
38 202.5 4.1 1938 1 S06005  
39 202 4.0 733 2 A46389  
40 201.5 4.0 1110 2 I51116  
41 201 4.0 475 2 S20250  
42 200.5 4.0 798 2 T33022  
43 200.5 4.0 1939 2 I48175  
44 199.5 4.0 1931 2 A59234  
45 198.5 4.0 303 2 S23780

#### ALIGNMENTS

##### RESULT 1

T18650  
hypothetical protein B0035.12 - Caenorhabditis elegans  
C:Species: Caenorhabditis elegans  
C:Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 15-Oct-1999  
C:Accession: T18650  
R:White, S.  
submitted to the EMBL Data Library, May 1996  
A:Reference number: Z19002  
A:Accession: T18650  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-836 <WIL>  
A:Cross-references: EMBL:Z73102; PIDN:CAA97405.1; GSPDB:GN00022; CESP:B0035.12  
A:Experimental source: clone B0035  
C:Genetics:  
A:Gene: CESP:B0035.12  
A:Map position: 4  
A:introns: 28/3; 78/2; 134/3; 167/1; 299/3; 349/2; 550/1; 662/3; 700/3; 799/1

Query Match 17.3%; Score 864; DB 2; Length 836;  
Best Local Similarity 27.7%; Pred. No. 1.7e-38;  
Matches 256; Conservative 185; Mismatches 366; Indels 118; Gaps 27;  
QY 65 SDGEYAMASSAESSESGEYEWYDEEEKNQLEIERLEQLSINVDYCHVDLRLRL 124  
DB 2 SDVD---MESGSDSGME-----DLDEETQIKQKMHDDSQSVVLAN-----QLLLLRK 48  
QY 125 EGELTKVRMAROKMSEIFPLTEELWLEWLD-----EISMAQDGLDREHVYDLPEKAVK 178  
DB 49 NGDFELDKRQFVEWAPLPLNKNWIEDFQNRKPEPSVAE-----VEEMFEKALF 101  
QY 179 DYICPNWLEYGOYSVGGIGQKGLK-----VRSVFERALSSVGLHMTKGLALWEAYREFE 235  
DB 102 DENDVTIWRAMAYKVKANDKKEDFKFCRDVCSKALENLGTRYDSSGHIWLIPEYE 161  
QY 236 SAIVEAA-----RL-EKVHSLFRQLAIPLYDMEATFAEYE-----EMSEDPPIPSVIQ 283  
DB 162 MSYLKNSMNPAYQRLADQCFALFALHCPDQLEDVYVLAEFQCFTEFKQHHKLEELKK 221  
QY 284 NYNKALQLEKYKPYEEALLQAPRAEYQAYIDFEMKIGDPARTIQLIFERALVENCLV 343  
DB 222 TYNSPMRQEQLSKEE-LIQOETKQGLKOFFDHEKSGIPSRKMAHERLVSELD 280  
QY 344 PDLWIRYGOYLRQKLVKDLVLSVHNRAINCPTWVALMSRYLLAMERHGVHQVTSVTF 403  
DB 281 EEAWIAYGAWADIELKLPQAVKVSRLRHCPSFVLHQALLAFERRRPNEEIDALW 340  
QY 404 EKALNAGFTQADYVEIWQAYLDYLRRRVDFQDSSKELEELRAAFTRALEYLKQVEER 463  
DB 341 ERARNSVINSABEGRSLRYTYAFLRRRIHL--TGSSDYSPMAEVEFDEGAALLR---EW 394

hypothetical prote  
myosin heavy chain  
ribonucleoprotein,  
hypothetical prote  
myosin heavy chain  
RNA-binding protei  
probable splicing  
myosin heavy chain  
myosin alpha heavy  
gene su(f) protein  
NF-180 - sea lamp  
splicing factor U2  
hypothetical prote  
myosin heavy chain  
slow myosin heavy  
nucleic acid-bind



QY 464 FNEGDPSCVIMONWARIERARLNNMOKARELWDSITWTRGNKAYANMWLEYYNLERAHGD 523  
Db 395 FSNMWTADYROMQAFYFASLMKMKCRNIWINDILASGFRGAFAGKWEAVRLERQFGD 454  
QY 524 TOHCRALHRAVQCTSDYEPHVCEVLLTWERTEGSLDWDIAVOKTETRLA-RVNEORMK 582  
Db 455 KENARKYLNALMSVSDNTNEIYMYVQFEREEGTLAEIDLVEKNSQVAHRAIRPQK 514  
QY 583 AAEKEAALVQOEKEAQRKAEKALKKIRKPEKRADEDEKEWGDDEEOP 642  
Db 515 VSEKAPAPKSDHQ--KRTSGGEPVKKV-----GDD----- 548  
QY 643 KRRRVENSIPAGETONVEVAGPAGKCAADVPEPSKOKAASLKRDMPKVLHDSSK 702  
Db 549 -----GGFKAPLPPSNKSSAVSSNASSTPAGSFAVQKAA-----PGTED 591  
QY 703 SITVFVSNLPYSMOEPDTKLRLPFEACGEVQIRPFSNRGD-----FRGYCYVEPEEKS 758  
Db 592 ARTIFVSNLDFTTTEDEIR-----QAIEGVASIR--FARKANSDLVHRGFAYVVMENDOK 644  
QY 759 ALQALEMDRKSVEGRPMFVSPCVDKSKNPDVKV-FRYSTSLKHLKLFISGLPFSCCTKEEL 817  
Db 645 AQOALLKDRVPVKGRPMFIS-----ANDPEKRVGFESTLEKSVFVNVHFQATDDDEL 699  
QY 818 EETCKAHGTVKDLRLVTNRAGPKGLAYVEYENESQAOAVMMQDGMTIKENIKVAISN 877  
Db 700 KALFSRFGVTVSVRRVTHDKGPKGIAFVDFOTEASAQKVASGDKMLRERELEVALSN 759  
QY 878 PPORVPEKPETRKAPGPMPLLPQTYGARGKGTQSLPRLALORPSAAPOANGPAA 937  
Db 760 PPVKK--DKSHGKPAAGASL--EEDGPRKGAHAKLQVPRATNKT---PQI-TARLDA 811  
QY 938 PAVAAPAATEAPKMSNADFAKLFLR 962  
Db 812 MDVSEGTSTSQP-LSNDQFRKFMK 835

RESULT 2  
T09890  
hypoetical protein T22A6.100 - Arabidopsis thaliana  
C:Species: Arabidopsis thaliana (mouse-ear cress)  
C:Date: 16-Jul-1999 #sequence\_revision 16-Jul-1999 #text\_change 15-Oct-1999  
C:Accession: T09890  
R:Bevan, M.; Zimmermann, W.; Gruenewald, A.; Wambutt, R.; Bancroft, I.; Mewes, H.W.; May  
submitted to the Protein Sequence Database, June 1999  
A:Reference number: Z16896  
A:Accession: T09890  
A:Molecule type: DNA  
A:Residues: 1768 <BEV>  
A:Cross-references: EMBL:AL078637; GSPDB:GN00062; ATSP:T22A6.100  
A:Experimental source: cultivar Columbia; BAC clone T22A6  
C:Genetics:  
A:Gene: ATSP:T22A6.100  
A:Map position: 4  
A:Introns: 71/3; 116/2; 134/3; 195/1; 265/3; 310/3; 352/3; 387/3; 435/2; 476/3; 501/3; 6

Query Match 16.5%; Score 823; DB 2; Length 768;  
Best Local Similarity 27.7%; Pred. No. 2.4e-36;  
Matches 264; Conservative 140; Mismatches 306; Indels 244; Gaps 36;

QY 56 DQOEGVSSDGDDEVAMASSASSPGEYEWEDDEEKNQLEIRLEQSLNVDYNCH 115  
Db 13 DQKMEDASAENPARADPPSSDSDS--GDSDSDEDAESNQ-QVITLESSELSANPYDAY 69  
QY 116 VDLIRLLLEGELTKVRMARQKMSFIFLPLELWLEWLHDEISMAQDGLDREHYVD---L 172  
Db 70 VQYIKLLRTANLEKLRQAREAMSAIFPLPSLWLEWARDEASLAAS-----ENVPEIVML 125  
QY 173 FEKAVXDYICPNWLEYGOY-----SVGGIGQKGLKVRVSRVFRALSSVGLHMTKGLA 226  
Db 126 YERGLSDYQSVSLWCMDYLSFMLEFDPSVRGYSE-GISKMSLIFERAIIPAAGFHVTEGR 184  
QY 227 LNEAYREFESAIVEARLEKVLHFRQLAIPLYDMEATFAEYEW-----SED 275

Db 185 IWEGYREFEQE-----NLSSTLIAYKTWELEQGLDLOGSD 221  
QY 276 --PIPESTQNTYKALQOEKYPYEEALLO--AEAPRAEYQAVIDEMKIGPPARIQ 330  
Db 222 LSKVSHQAVAVANKAQOQYSERAHLEENISKODLSDEKFPQFMNVIKEKTSQGDPTRVQ 281  
QY 331 LIFERALVENCULVPDLWIRYSOYLDORQLKVKDVLVSHNRARINCPWTVALMSRYLLANE 390  
Db 282 AIYERAVAEYPPVSSDLWIDYTVYLDKTLKVGKATHAYSEATRSCTPWTGDLWARYLLALE 341  
QY 391 RHGVHQVLSVTFEKALNAGFTQADYVIMQAYLDYLRVRVDFKODSKEELEELRAAPT 450  
Db 342 RGSASEKET-----YDYLTLRLTRVDGLRRM-----LSTRML 374  
QY 451 RALEYLKQVEBERFENSGDPSVIMON-----WARIEARLNNMOKARELWDSITM 500  
Db 375 EALDY--SLIRETFOQASDYLPHMONTSLHLHTYWANLEINLTKGLUAGARGVWDSPL 432  
QY 501 TRGNKAYANMWLEYYNLERAHGDTQCRKALHRAVQCTSD--YPERHVCVLLTMBTEGS 558  
Db 433 KKSGLMLA-AWHAYIDMEVHLGHKEARSIIYRCYTRKFDGTGSEDICKGWLRFEREHGD 491  
QY 559 LEDWDIAVOKTETRLARVNEQRMKAEEKALVQOEKEAQRKAEKALKKMKKTR 618  
Db 492 LEHFDLAVQKVMPL-----EELQMLRLQOQESTPVKPSAGLKEHSSOKRK--- 536  
QY 619 GPEKRGADDEDEKENGDEEOPSKR--RRVENSIPAAGE-----TONVEVAAGPAGK 670  
Db 537 -----AEO-----NVEESLAKRQKRKSQKQKQVLDGQGSATVATRNK---AENGK 580  
QY 671 AADVPEPSKQKEAASLKRDMPKVLHDSSKSDITVFNLSYMOEPDTKLRLPFEACG 730  
Db 581 ADSDK-----ETEDAKPLK---PKVY---RDECTAFISNLSVKAQED--IRKFFGDDG 627  
QY 731 EVQVIRPI-FSNRGDFRGYCYVEPEEKEKSAQAQALEMDRKSVEGRPMFVSPCVDKSKNPD 789  
Db 628 GVDISIRILHHKDTGPRGLAYADFDVDEHLAAIAAKRMKMFEGKISIA-----RSNP-- 680  
QY 790 KVFYSTSLKHLKLFISGLPFSCCTKEELEICKAHGTVDKDLRLVTNRAGPKGLAYVEY 849  
Db 681 -----KKCK-----KEFTRGNDGSGNSKQDPSLISEKAKAPLG----- 713  
QY 850 NESQASQAVMMQDGMTIKENIKVAISNPPQRPVPEKPEKTRKAPGPMPLLPQTYGARGK 909  
Db 714 GETEGER-----KGNEVEV-----RGR- 730  
QY 910 RTQLSLPRLALORPSAAPOANGPAAAPAAVAPAAATEAPKMSNADFAKLFLRK 963  
Db 731 --NTFAVPRNVKAPLGYYTPK-----PSADETPK-SNDEFRNMFLKK 768

RESULT 3  
T39743  
U4706 splicing factor PRP24 homolog - fission yeast (Schizosaccharomyces pombe)  
C:Species: Schizosaccharomyces pombe  
C:Date: 03-Dec-1999 #sequence\_revision 03-Dec-1999 #text\_change 07-Dec-1999  
C:Accession: T39743  
R:Oliver, K.; Harris, D.; Lyne, M.H.; Rajandream, M.A.; Barrell, B.G.  
submitted to the EMBL Data Library, August 1999  
A:Reference number: Z21876  
A:Accession: T39743  
A:Status: preliminary; translated from GB/EMBL/DBDJ  
A:Molecule type: DNA  
A:Residues: 1-1014 <OLI>  
A:Cross-references: EMBL:AL109834; NID:e1536510; PIDN:CAB52740.1; GSPDB:GN00067; SPDB  
A:Experimental source: strain 972h; cosmid c1861  
C:Genetics:  
A:Gene: SPDB:SPBC1861.04c  
A:Map position: 2  
A:Introns: 78/1; 166/3

Query Match 10.1%; Score 506.5; DB 2; Length 1014;

Best Local Similarity 22.2%; Pred. NO. 2.5e-19;

Matches 195; Conservative 163; Mismatches 381; Indels 139; Gaps 30;

```
QY 82 EWEYDEEENKQLEIEERLEBOLSINYDYNCHVDLIRLLECELETKVMAQKMEI 141
Db 13 EMDIDPQONEKPLLDDELTKFTLLHPYNDYSHIKLEELRDLKKESEARFTFSI 72
QY 142 PPLTEELWLEHDEISMAQDGLDREHYVDYDFEAKVDYICPNWLEYQYSGVGIGQG 201
Db 73 PPLSEDLAVDLLDECKNCRLLDDVVRKTLFDLAVQDYLKIKWCMYLEFTLNLMDTSS 132
QY 202 -----GLEKVSVERFALSSVGLHMTKGLALWEAYRE-----FESAIYAAKLE 245
Db 133 FEQEQSELNGVITLTDHAHSLERAYQTCFHFESQCVMTLYLEPFTGDALEFGESEQ 192
QY 246 -----KVHSLFRRLALPLVDMETFAEY-----EWSDEPIPESVIONYKALQOLEK 294
Db 193 ELTVFNKYYDHIHDKLPHQIEETFSLSFTVTNNWSPSEY-EDVMVKSNKYVETTLK 251
QY 295 --YKPYEALLQAEAPLIAEYQAYIDFEMKIGDPAR-----IQLIFERALVENCILVP 344
Db 252 RNAKIFNKELLNSANSLSLE--AYMDL---INDESRRSTAELOQYITTLVRAIVLYPLIP 306
QY 345 DLWIRYSOYLORQLKVLIVLSVHNRAIRNCPWTVALMSRYLLAMERHGVHQVTSVTFE 404
Db 307 ELWLQYTAWLKVDPSQSSQASSVAERATRNCSWIGRIWSIKLITYMTLSGASTSAVCEEKD 366
QY 405 KALNAGFIQATDYVEIQWYLDYLRVRVDKODSSKELEELRAATRALEYLKQEVVERF 464
Db 367 RCLNSNLL--VNFDEVIDFFGFLKACLYLSNEDKPOEFLKHQTHKVEDYLRK----- 418
QY 465 NESGDPSCVIMONWARIEARL-----CNMOKARELWDSIMTRNAKYANMWLE--- 513
Db 419 NHKSKD-----ARMIELSKIYLSIISDFESVEKWC-SDMFDHFNQALYISRYI 470
QY 514 ---YNNLERAGDTHQKRALHRAVOCTSDYPEHVCEVLLMTERTEGSLDWDIAVQKTE 570
Db 471 STMKYNPELA---AETLKSKLYKNV---DOPO-----LIFQFYQSIMDLNDCFTNTS 517
QY 571 TRLARVNQRM---KAAEKEALVQEEKABQRARAEEKALKKKKIRGPEKRGAD 627
Db 518 HLYVDLNAQRTISFKQLQSFAETKTQVTEPLEKVPQADDTAALSRRKKGQEQ----- 571
QY 628 DDEKEWGD-DEEOPSKRRR-----VENSIPAAGETQNVVEAAGPAGKCAADVPEPS 679
Db 572 -----GDVEKSKPIEQHRNEELTVLTVLPSDISNELKIPFKDCGNIIRIILEDN 625
QY 680 KQKKAASLK-----RDMPKVL-HDSSKD---SITVTVSNLPSYMOEPDTKLR 723
Db 626 QKDQVQAQTEFSETSEVLAATKTRDLKSIRGHEISVQIHVDNTNLYTNFPPTDELD--IT 683
QY 724 PLFEACGEVVOIRPISNRGDPFGYCYVEFEKESALQALEMDRKSVEGR---PMFVSPC 780
Db 684 KLFSAYGNNVDVFRFSLRYNTNRRFCYQVMRPDPDAHNALQLHKKLLEKYPIQVIFIS-- 741
QY 781 VDKSKNPDKVFRYSTLSKHKLIFSLGPFSCFCKEELEICKAHGTVKDLRLVTRNAGKP 840
Db 742 -----DPLRRTPRSGAVVEGRELYVTNIDFKVNEKDVETFFRDYQGVESVR-IPKRFNOH 795
QY 841 KGLAVYENESQASQAVKMDGWTIKENIKVAISNP 878
Db 796 KGFYGVVMTTNOAENA-LSAAGKOLGNRVNLVLSKP 832
```

## RESULT 4

S18874

nucleolin - African clawed frog

C.Species: Xenopus laevis (African clawed frog)

C.Date: 06-Jan-1995 #sequence\_revision 06-Jan-1995 #text\_change 23-Jul-1999

C.Accession: S30250; A30166; S18874

R:Rankin, M.L.; Heine, M.A.; Xiao, S.; LeBlanc, M.D.; Nelson, J.W.; DiMario, P.J.

Nucleic Acids Res. 21, 169, 1993

A.Title: A complete nucleolin cDNA sequence from Xenopus laevis.

A:Reference number: S30250; MUID:931811171; PMID:8441611

A:Accession: S30250

A:Status: preliminary

A.Molecule type: mRNA

A.Residues: 1-651 <RAN>

A:Cross-references: EMBL:X63091; NID:g64936; PIDN:CAA44805.1; PID:g64937

R:Caizergues-Ferrer, M.; Mariottini, P.; Curie, C.; Lapeyre, B.; Gas, N.; Amalric, F.

Genes Dev. 3, 324-333, 1989

A>Title: Nucleolin from Xenopus laevis: cDNA cloning and expression during development

A:Reference number: A30166; MUID:89252811; PMID:2656405

A:Accession: A30166

A:Status: not compared with conceptual translation

A.Molecule type: mRNA

A.Residues: 126-214, 'Q', 216-218, 'LR', 221-410, 'Q', 412-580, 'E', 582-651 <CAT>

C:Superfamily: nucleolin; ribonucleoprotein repeat homology

C:Keywords: DNA binding; nucleus; phosphoprotein

F:234-299/Domain: ribonucleoprotein repeat homology <RRM1>

F:326-389/Domain: ribonucleoprotein repeat homology <RRM2>

F:416-478/Domain: ribonucleoprotein repeat homology <RRM3>

F:504-568/Domain: ribonucleoprotein repeat homology <RRM4>

Query Match 6.1%; Score 305; DB 2; Length 651;

Best Local Similarity 26.5%; Pred. No. 7.7e-09;

Matches 104; Conservative 70; Mismatches 145; Indels 74; Gaps 14;

```
QY 552 MERTEGSLDWDIAVQKTKETRLARVNEQRM--KAAE-----KEAALVQEEEEK 597
Db 34 MEEDSSDEEVEVPVKTKTATPAKATPGKATPGKKGATPAKNGKQAKKQSEEEEE 93
```

```
QY 598 AQQRARAERKALKK---KKKIQPEKRGADDEDEKEMGDDEEQ-----PS 642
Db 94 DSDDEAEQDKPIKNKPVAKKAVAKKESEEDDDDEDE---SEEEKAVAKKTPAKKPAG 150
```

```
QY 643 KRRRVENSIPAAGETQNVVEAAGPAGKCAAV-----DVEPPSKQKEKAASLKR 690
Db 151 KQSESEEDDEESEPMEVAPALGKKTAAQAAEEDDEEDDDDEDEEQQSAKR 210
```

```
QY 691 --DMPKVL-----HDSSKDSITVFSNLPYSMOEPDTK--LRPLFEACGEVVOIRP 739
Db 211 KEMPKTPEAKTKTDTASEGLSIFIGNLSTKEFDELKDALREFFSKNLTQIDIRI- 269
```

```
QY 740 SNRGDFRGYCYVEFEKESALQALEMDRKSVEGRPMFVSPCV--DKSKNPDKVFRYSTS 797
Db 270 ---GSKKFGYVDFSEEEVEKALKLTGKKILGTVEKIEKAMAFDKNKTAEKKER---- 322
```

```
QY 798 LEKHLIFSLGPFSCFCKEELEICKAHGTVKDLRLVTRNAGKPGKGLAVYVEYENESQAS 857
Db 323 -DSRRTLFAKNIDPYSITVVEELQEIFE---NAKDIRIPTGKDSNGKIAYVEFSNEDANKA 378
```

```
QY 858 VNMKGMTIKENIKVAIS-----NPPQRKVPE 885
Db 379 LEEKQGAETEGRSIFVDFTEGKSQNSGNNKKGPE 411
```

## RESULT 5

S32644

nucleolin - African clawed frog

C.Species: Xenopus laevis (African clawed frog)

C.Date: 06-Jan-1995 #sequence\_revision 06-Jan-1995 #text\_change 23-Jul-1999

C.Accession: I51619; S32644

R:Messmer, B.; Dreyer, C.

Eur. J. Cell Biol. 61, 369-382, 1993

A>Title: Requirements for nuclear translocation and nucleolar accumulation of nucleolin

A:Reference number: I51619; MUID:94039235; PMID:8223724

A:Accession: I51619

A:Status: preliminary; translated from GB/EMBL/DBJ

A.Molecule type: mRNA

A.Residues: 1-705 <ME2>

A:Cross-references: EMBL:X72957; NID:g295898; PIDN:CAA51460.1; PID:g295899

C:Superfamily: nucleolin; ribonucleoprotein repeat homology

F:285-350/Domain: ribonucleoprotein repeat homology <RRM1>

F:377-440/Domain: ribonucleoprotein repeat homology <RRM2>

F:467-529/Domain: ribonucleoprotein repeat homology <RRM3>

Db 78 PDEI



	610	ALUKKKKTRGPPKRGAGDEDEKENGDDDEEQFSRRRVENSIPAAGTQNVVEAAGPAGK	6699
Qy			
	209	ATPAKAABESEDDEDEDEDDEDEDESEDEKPVKEAPGRKRKNKMSAPE	2688
Db			
	670	CAAVNVEPPSK-----OKFAASIKR-----	690
Qy			

DB, 269 AKKKKTETPASAFSLFVKNLTPTRKDYEEELRTAIKEFFGKKNLQVSEVRIGSSKRFFGYVDF 328

```

991      QY       691    ----DMPKVL-----HDSKSDSTTFVSNLPYSMQPDDTK 721
           |||||               :|: |:||| ||| :|:
329      Db     LSAAEDMKAOLUNGLKKLMGLEIKLKAKSKESLKENKERDARTLFVKNLPLYRTVEDE-- 386
           :|:|:|              :|:|:|         :|:|:|
722      QY     LRLPEACGEVVOIRPIFSNRGDFRGVCYVFEEKEKSALOALEMDR-KSVEGRPMFVSPC 780
           ::::|:|:|          :|:|:|         :|:|:|         :|:|:|
387      Db     MKNVFE---NALEVLVLVNKGSGSKGMAYIEFKTEAEAEKALBEQGTVDVGRAWIDYT 443
           :|:|:|             :|:|:|         :|:|:|         :|:|:|
781      QY     VDKSNPDPKFYRYSTSLEKHKLFIISGLPFSCCTKEELEECIAHGHTVVKDLURLVTNRRAGKP 840
           :|:|:|             :|:|:|         :|:|:|         :|:|:|
444      Db     GEKSOEQSQS---GGGERESTKLIVNNLSVAASEETQLQELFKKATSIK---MPQNNGGRP 497
           :|:|:|             :|:|:|         :|:|:|         :|:|:|
841      QY     KGLAVYEVENESOASQAQVMKMGDGMTIRENIITKVAINPPORKVPKPETRKAPGPMLLP 900
           |||:|:|:|         :|:|:|         :|:|:|         :|:|:|
498      Db     KYIAYFEPPTADAKEALNSCNTTEIGRAIRLEFFSPWQK-----GNMNARGGFNQOS 552
           :|:|:|             :|:|:|         :|:|:|         :|:|:|
901      QY     QTYGARG 907
           :|:|
553      Db     KTLFVRG 559
           :|:|

RESULT 10
A35804
nucleolin - human
N:Alternate names: phosphoprotein ppl00; protein B50; protein C23
C:Species: Homo sapiens (man)
C>Date: 23-Oct-1990 #sequence-revision 123-Oct-1990 #text_change 23-Jul-1999
C:Accession: A35804; S04631; A48138; A55996
R:Srivastava, M.A.; McBride, O.W.; Fleming, P.J.; Pollard, H.B.; Burns, A.L.
J. Biol. Chem. 265, 14922-14931, 1990
A:Title: Genomic organization and chromosomal localization of the human nucleolin gene
A:Reference number: A35804; MUID:90368666; PMID:2394707
A:Accession: A35804
A:Molecule type: DNA
A:Residues: 1-707 <SR>
A:Cross-references: GB:M60858; GB:J05584; NID:g189305; PIDN:AAA59954.1; PID:g189305
R:Srivastava, M.; Fleming, P.J.; Pollard, H.B.; Burns, A.L.
FEBS Lett. 250, 99-105, 1989
A:Title: Cloning and sequencing of the human nucleolin cDNA.
A:Reference number: S04631; MUID:89290043; PMID:2737305
A:Accession: S04631
A>Status: not compared with conceptual translation
A:Molecule type: mRNA
A:Residues: 1-707 <SR2>
R:Ishikawa, F.; Macunis, M.J.; Dreyfuss, G.; Cech, T.R.
Mol. Cell. Biol. 13, 4301-4310, 1993
A:Title: Nuclear proteins that bind the pre-mRNA 3' splice site sequence r(UUAGAUGUUUA).
A:Reference number: A48138; MUID:93309464; PMID:8321232
A:Accession: A48138
A:Molecule type: protein
A:Residues: 458-474 <ISH>
A:Experimental source: HeLa cell nuclei
R:Nordan, P.; Heid, H.; Kinzel, V.; Kuebler, D.
Biochemistry 33, 14696-14706, 1994
A:Title: Major cell surface-located protein substrates of an ecto-protein kinase
```

A:Reference number: A55996; MUID:95080603; PMID:7993898  
A:Accession: A5596  
A:Molecule type: protein  
A:Residues: 231-236;349-362;399-403;458-461;655-656, 'x', 658-660 <JOR>  
A:Experimental source: surface-labelled HeLa cells  
C:Genetics:  
A:Gene: GDB:NCL  
A:Cross-references: GDB:125908; OMIM:164035

A;Gene: GDB:NCL  
A;Cross-references: GDB:125908; OMIM:164035









**THIS PAGE BLANK (USPTO)**  
**BEST AVAILABLE COPY**

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 18, 2003, 13:46:53 ; Search time 15 Seconds  
(without alignments)  
2662.780 Million cell updates/sec

Title: US-09-763-985A-2  
Perfect score: 4994  
Sequence: 1 MATAAETSASEPEAESKAGP.....AATEAPKMSNADFAKFLRK 963

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt\_40:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	305	6.1	650	1 NUCLEO_XENLA	P20397 xenopus lae
2	272	5.4	414	1 NSRL_YEAST	P27476 saccharomyc
3	268.5	5.4	702	1 CRN_DROME	P17886 drosophila
4	262	5.2	694	1 NUCLEO_CHICK	P15771 gallus gall
5	245.5	4.9	706	1 NUCLEO_HUMAN	P19338 homo sapien
6	244.5	4.9	706	1 NUCLEO_MOUSE	P09405 mus musculu
7	238.5	4.8	960	1 K882_HUMAN	Q9Y4C8 homo sapien
8	225	4.5	519	1 ELAV_DROVI	P23241 drosophila
9	224.5	4.5	713	1 NUCLEO_MESAU	P08199 mesocricetu
10	224	4.5	712	1 NUCLEO_RAT	P13383 rattus norv
11	220.5	4.4	1938	1 MYHD_HUMAN	Q9UKX3 homo sapien
12	217.5	4.4	1972	1 MYHD_MOUSE	O08638 mus musculu
13	216	4.3	1130	1 YL17_CAEEL	Q11102 caenorhabdi
14	211.5	4.2	500	1 GAR2_SCHPO	P41891 schizosacch
15	210	4.2	483	1 ELAV_DROME	P16914 drosophila
16	209	4.2	978	1 RA50_AQUAE	O67124 aquifex aeo
17	208.5	4.2	496	1 U2AF_CAEEL	P09078 caenorhabdi
18	208.5	4.2	1972	1 MYHB_HUMAN	P35749 homo sapien
19	207	4.1	1959	1 MYH9_CHICK	P14105 gallus gall
20	206	4.1	315	1 ROC4_NICSY	P19683 nicotiana s
21	205.5	4.1	1972	1 MYHB_RABIT	P35748 oryctolagus
22	204	4.1	475	1 U2AF_HUMAN	P26368 homo sapien
23	204	4.1	475	1 U2AF_MOUSE	P26369 mus musculu
24	203.5	4.1	1937	1 MYH8_HUMAN	P13535 homo sapien
25	202.5	4.1	1938	1 MYH8_RAT	P02563 rattus norv
26	202	4.0	733	1 SUF_DROME	P25991 drosophila
27	200.5	4.0	1939	1 MYH6_MESAU	P13539 mesocricetu
28	199	4.0	488	1 U2AF_CAEER	P09727 caenorhabdi
29	197.5	4.0	845	1 SCPL1_MESAU	Q60563 mesocricetu
30	197.5	4.0	1935	1 MYSS_CYPCA	Q90339 cyprinus ca
31	195.5	3.9	1935	1 MYH7_HUMAN	P12883 homo sapien
32	194.5	3.9	674	1 CMF4_SCHPO	P87312 schizosacch
33	193	3.9	1934	1 MYH7_MESAU	P13540 mesocricetu

## RESULT 1

ID	NUCL_XENLA	STANDARD;	PRT;	650 AA.
AC	P20397;			
DT	01-FEB-1991 (Rel. 17, Created)			
DT	01-FEB-1994 (Rel. 28, Last sequence update)			
DT	15-JUN-2002 (Rel. 41, Last annotation update)			
DE	Nucleolin (Protein C23).			
OS	Xenopus laevis (African clawed frog).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae;			
OC	Xenopodinae; Xenopus.			
OX	NCBI_TaxID=8355;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	TISSUE=Ovary;			
RX	MEDLINE=93181171; PubMed=8441611;			
RA	Rankin M.L., Helne M.A., Xiao S., Leblanc M.D., Nelson J.W.,			
RA	Dimario P.J.;			
RT	"A complete nucleolin cDNA sequence from Xenopus laevis.";			
RL	Nucleic Acids Res. 21:169-169(1993).			
RN	[2]			
RP	SEQUENCE OF 125-650 FROM N.A.			
RX	MEDLINE=89252811; PubMed=2656405;			
RA	Caizergues-Ferrer M., Mariottini P., Curie C., Lapeyre B., Gas N.,			
RA	Amalric F., Amaldi F.;			
RT	"Nucleolin from Xenopus laevis: cDNA cloning and expression during			
RT	development.";			
RL	Genes Dev. 3:324-333(1989).			
CC	FUNCTION: NUCLEOLIN IS THE MAJOR NUCLEOLAR PROTEIN OF GROWING			
CC	EUKARYOTIC CELLS. IT IS FOUND ASSOCIATED WITH INTRANUCLEOLAR			
CC	CHROMATIN AND PRE-RIBOSOMAL PARTICLES. IT INDUCES CHROMATIN			
CC	CONDENSATION BY BINDING TO HISTONE H1. IT IS THOUGHT TO PLAY A			
CC	ROLE IN PRE-RNA TRANSCRIPTION AND RIBOSOME ASSEMBLY.			
CC	CELLULAR LOCATION: Nuclear; nucleolar.			
CC	SIMILARITY: CONTAINS 4 RNA RECOGNITION MOTIFS (RRM).			
CC	THIS SWISS-PROT entry is copyright. It is produced through a collaboration			
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation			
CC	at the European Bioinformatics Institute. There are no restrictions on its			
CC	use by non-profit institutions as long as its content is in no way			
CC	modified and this statement is not removed. Usage by and for commercial			
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/			
CC	or send an email to license@isb-sib.ch).			
CC	EMBL; X63091; CAA44805.1;			
DR	PIR; A30166; A30166.			
DR	PIR; S18874; S18874.			
DR	PIR; S30250; S30250.			
DR	HSSP; P11940; ICWJ.			
DR	InterPro; IPR000504; RNA_rec_mot.			
DR	Pfam; PF00076; rrm; 4.			
DR	SMART; SM00360; RRM; 4.			
DR	PROSITE; PS50102; RRM; 4.			
DR	PROSITE; PS00030; RRM_RNP_1; 3.			
KW	Nuclear protein; Phosphorylation; Methylation; DNA-binding; Repeat;			

34	192.5	3.9	329	1 ROC3_ARATH	Q04836 arabidopsis
35	191.5	3.8	1020	1 NFH_HUMAN	P12036 homo sapien
36	191	3.8	4684	1 PLE1_HUMAN	Q15149 homo sapien
37	190.5	3.8	976	1 SCPL1_HUMAN	Q15431 homo sapien
38	189	3.8	1940	1 MYH3_RAT	P12847 rattus norv
39	189	3.8	1961	1 MYH9_RAT	Q62812 rattus norv
40	188.5	3.8	1325	1 G160_MOUSE	P55937 mus musculu
41	188	3.8	233	1 ROC1_SPIOL	P28644 spinacia ol
42	187.5	3.8	1938	1 MYH6_MOUSE	Q02566 mus musculu
43	187.5	3.8	1960	1 MYH9_HUMAN	P35579 homo sapien
44	187.5	3.8	3321	1 PCN2_HUMAN	O95613 homo sapien
45	187	3.7	1935	1 MYH7_RAT	P02564 rattus norv

## ALIGNMENTS

STRAIN-S288c;  
MEDLINE=92355583; PubMed=1644811;  
Kondo K., Inouye M.;  
"Yeast NSRI protein that has structural similarity to mammalian  
nucleolin is involved in pre-rRNA processing.";  
J. Biol. Chem. 267:16252-16258(1992).  
[3]  
SEQUENCE FROM N.A.  
RN RP  
STRAIN-S288c;  
MEDLINE=96158062; PubMed=8585325;  
Skala J., Nawrocki A., Goffeau A.;  
"The sequence of a 27 kb segment on the right arm of chromosome VII  
from Saccharomyces cerevisiae reveals MOL1, NAT2, RPL30B, KSRL, CYS4,  
PEM1/CHO2, NSRI genes and ten new open reading frames.";  
Yeast 11:1421-1427(1995).  
[4]  
SEQUENCE OF 188-414 FROM N.A.  
RN RP  
Rieger M., Mueller-Auer S., Brueckner M., Schaefer M.;  
Submitted (MAY-1996) to the EMBL/GenBank/DDBJ databases.  
[5]  
RNA-BINDING.  
RN RP  
MEDLINE=95098604; PubMed=7800479;  
Lin J.-J., Zakian V.A.;  
"Isolation and characterization of two Saccharomyces cerevisiae genes  
that encode proteins that bind to (TGI-3)n single strand telomeric  
DNA in vitro.";  
Nucleic Acids Res. 22:4906-4913(1994).  
RN RL  
-!- FUNCTION: INVOLVED IN PRE-RRNA PROCESSING. SPECIFICALLY BINDS  
NUCLEAR LOCALIZATION SEQUENCES. CANDIDATE FOR A RECEPTOR AT THE  
NUCLEUS THAT MAY BE INVOLVED IN BOTH RNA AND PROTEIN TRANSPORT.  
BINDS TELOMERIC SEQUENCES OF THE TYPE (TG[1-3])n IN VITRO.  
RN CC  
-!- SUBCELLULAR LOCATION: NUCLEAR; POSSIBLY AT THE NUCLEOLUS.  
RN CC  
-!- INDUCTION: IN RESPONSE TO LOW TEMPERATURE (BY COLD-SHOCK).  
RN CC  
-!- SIMILARITY: BELONGS TO THE GAR FAMILY.  
RN CC  
-!- SIMILARITY: CONTAINS 2 RNA RECOGNITION MOTIFS (RRM).  
-----  
This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See http://www.isb-sib.ch/announce/  
or send an email to license@isb-sib.ch).  
-----  
CC CC  
EMBL; X57185; CAA40472.1; -  
DR DR  
EMBL; X85807; CAA59817.1; -  
DR DR  
EMBL; Z72944; CAA97173.1; -  
DR DR  
EMBL; Z72946; CAA97180.1; -

```

SGD: SU003391; NSKL.
InterPro; IPR000504; RNA_rec_mot.
Pfam; PF00076; Rrm; 2.
SMART; SM00360; RRM; 2.
PROSITE; PS50102; RRM; 2.
PROSITE; PS60030; RRM_RNP_1; 2.
DNA-binding; RNA-binding; Nuclear protein; Repeat; rRNA processing.
DOMAIN 29 136 SR/ASP/GLU-RICH.
FT DOMAIN 168 246 RNA-BINDING (RRM) 1.
FT DOMAIN 267 345 RNA-BINDING (RRM) 2.
FT DOMAIN 356 384 RNA-BINDING RGG-BOX (BY SIMILARITY).
SEQUENCE 414 AA; 44535 MW; 90DEEE7BBC20BC0C CRC64;

Query Match          5.4%; Score 272; DB 1: Length 414;
Best Local Similarity 22.2%; Pred.No.9.5e-08;
Matches 84; Conservative 84; Mismatches 172; Indels 38; Gaps 9;

554 RTEGSLDWIDIAVQKTTFLARVNEGRMAAEKAALVQOEKEEKAKQRKRAR----- 605
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
      6 KYGNKKVEVKASQAKEEAKAVSSSSSSSSSSSESESESESESSSSSSSDSE 65

606 -----AEKKALKKKKKIRGEKRGADDDKEKWGDDEEQSPSKRRRVENSIPAAGT 657
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
      66 SSSSSSSDSSEAATYKKESKQSSSSSSSDSSDEE-----EEEEKETKEEKESSSDS 121
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :

```





```
Db 149 AAKSAAVPAKPAVVPKQSESEEEDEDEDEDESEDEAMDTTPAPVKPAPAK 208
Qy 610 ALKKKKKIRGKRGADDEKWDDEEOPSKRRRVENSIPAAGETQNVVEAAGPAGK 669
Db 209 ATPAKAAESEDDEDEDEDEDEDEDEDEDEDEKVPKAPGKRKKEMANKSAPE 268
Qy 670 CAAYDVPEPSK-----QKKAASLKR----- 690
Db 269 AKKKKTETPASAFSLFVKNLTPTKDYBELRTAIKEFTGKKNLQVSEVRIGSKRGYVDF 328
Qy 691 -----DMPKVL-----HSSKDSITVFSNLSYMOEPTDK 721
Db 329 LSAEDMDKALQNGKLMGLKLEKAKSKESLENKKERDARTLFVKNLPRVYDE-- 386
Qy 722 LRPLFEAGGVVQIRPFIENSGDFGYCYVEEKSALQALEMDR-KSVGEWFEVSPC 780
Db 387 MKNVE---NALEVLVLNKGSSGKGMAYIEFKTEAEAKALEKQGTEDVGRAMIDYT 443
Qy 781 VDSKNPDFKVFYRSTSLKHKLFISGLPFSCFKEELEICKAHGTVDLRLVNRAGKP 840
Db 444 GEKSQESQK---GGGERESKTLIVNLSYAASETLQELFKKATSIK---MPONNOGRP 497
Qy 841 KGLAYVEYENESQASQAVMKMDGMTIKENIKVAISNPPQKRVPEKPTRKAPGGMILLP 900
Db 498 KGYAFVEFPTAEDAKEALNSCNTEIEGRAIRLEFSPSWQK-----GNMARGGFNQQS 552
Qy 901 QTYGARG 907
Db 553 KTLFVRG 559
```

## RESULT 5

```
NUCL_HUMAN STANDARD; PRT; 706 AA.
ID AC P19338;
DT 01-NOV-1990 (Rel. 16, Created)
DT 01-NOV-1990 (Rel. 16, Last sequence update)
DE 15-JUN-2002 (Rel. 41, Last annotation update)
DE Nucleolin (Protein C23).
GN NCL.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=89290043; PubMed=2737305;
RA Srivastava M., Fleming P.J., Pollard H.B., Burns A.L.;
RT "Cloning and sequencing of the human nucleolin cDNA.";
RL FEBS Lett. 250:99-105(1989).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=90368666; PubMed=2394707;
RA Srivastava M., McBride O.W., Fleming P.J., Pollard H.B., Burns A.L.;
RT "Genomic organization and chromosomal localization of the human nucleolin gene.";
RL J. Biol. Chem. 265:14922-14931(1990).
CC -!- FUNCTION: NUCLEOLIN IS THE MAJOR NUCLEOLAR PROTEIN OF GROWING
CC EUKARYOTIC CELLS. IT IS FOUND ASSOCIATED WITH INTRANUCLEOLAR
CC CHROMATIN AND PRERIBOSOMAL PARTICLES. IT INDUCES CHROMATIN
CC DECONDENSATION BY BINDING TO HISTONE H1. IT IS THOUGHT TO PLAY A
CC ROLE IN PRE-RRNA TRANSCRIPTION AND RIBOSOME ASSEMBLY.
CC -!- SUBCELLULAR LOCATION: Nuclear; nucleolar.
CC -!- SIMILARITY: CONTAINS 4 RNA RECOGNITION MOTIFS (RRM).
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/
CC or send an email to license@isb-sib.ch).
```

```
CC ----- M60858; AAA59954.1; -----
DR EMBL; A35804; A35804.
DR PIR; S04631; S04631.
DR HSSP; P09651; IHA1.
DR Aarhus/Ghent-2DPAGE; 1210; NEPHGE.
DR Genew; HGNC:7667; NCL.
DR MIN; 164035;
DR InterPro; IPR000504; RNA_rec_mot.
DR Pfam; PF00076; rrm; 4.
DR SMART; SM00360; RRM; 4.
DR PROSITE; PS01012; RRM; 4.
DR PROSITE; PS00303; RRM_RNP_1; 3.
KW Nuclear protein; Phosphorylation; Methylation; DNA-binding; Repeat;
KW RNA-binding.
FT INIT_MET 0 0
FT DOMAIN 142 170 ASP/GLU-RICH (ACIDIC).
FT DOMAIN 184 208 ASP/GLU-RICH (ACIDIC).
FT DOMAIN 233 270 ASP/GLU-RICH (ACIDIC).
FT DOMAIN 306 382 RNA-BINDING (RRM) 1.
FT DOMAIN 392 465 RNA-BINDING (RRM) 2.
FT DOMAIN 485 559 RNA-BINDING (RRM) 3.
FT DOMAIN 571 643 RNA-BINDING (RRM) 4.
FT DOMAIN 645 694 ARG/GLY/PHE-RICH.
FT DOMAIN 57 134 8 X 8 AA TANDEM REPEATS OF X-T-P-X-K-K-
FT REPEAT 57 64 X-X.
FT REPEAT 74 81 1.
FT REPEAT 82 89 2.
FT REPEAT 90 97 3.
FT REPEAT 98 103 4.
FT REPEAT 104 111 5 (INCOMPLETE).
FT REPEAT 119 126 6.
FT REPEAT 127 134 7.
FT MOD_RES 144 144 PHOSPHORYLATION (BY SIMILARITY).
FT MOD_RES 152 152 PHOSPHORYLATION (BY SIMILARITY).
FT MOD_RES 183 183 PHOSPHORYLATION (BY SIMILARITY).
SQ SEQUENCE 706 AA; 76213 MW; 85A2F2CA22EA03DB CRC64;
Query Match 4.9%; Score 245.5; DB 1; Length 706;
Best Local Similarity 24.2%; Pred. No. 5e-06;
Matches 92; Conservative 64; Mismatches 127; Indels 97; Gaps 14;
Qy 595 EEKAPQRKRAEAKKA-----LKKKKKIRGPEKRGAD-----DEKEMGD 635
Db 206 EEEAMETTPAKGKKAAPVYVAKNVAEDEDEDEDEDEDEDEDEDEDEDEDEDE 265
Qy 636 DEEQP-----SKRRRVENSIP-----KRDMPKVLH 697
Db 266 EEEEPVKEAPGKRKKEMAKKAAPEAKKQKVEGTEPTAFNLFVGNLNFNKSAPELKTG 325
Qy 653 -----AAGETQNVAAAGPAGKCAAVDVEPFSKQKEKAAL-----KRDMPKVLH 697
Db 326 ISDVFAKNDLAVVDVRIGMTKRFYGVDFE-SAEDELEKALELTGLVKVFNGLKPKG-K 383
Qy 698 DSSK--DSITVFVSNLPSMOBPTDKLRPLFEACGEVQVQIRPFIENSGDFGYCYVEFKE 755
Db 384 DSKKEDARTLLAKNLKYVTQDE--LKEVFDAAEI----RLVSKDGSKSGIAIEFKT 437
Qy 756 EKSAQLAQLAEMDR-KSVEGRPMFVSPCVDKSKNPDKVFRYST-SLEKHKLFISGLPFSC 813
Db 438 EADAETFEKQGTEDIGRSISLYVTGEKQNDYRGKKNSTWSGESKTLVLSNLSYSAT 497
Qy 814 KEELEICKAHGTVDLRLVNRAGKPKGLAYVEYENESQASQAVMKMDGMTIKENIKV 873
Db 498 EETLQEVFEKATFIK---VPQNGKSGYAFIEFASFEDAKEALNSCNKREIEGRAIRL 554
Qy 874 AISNPPQKRVPEKPTRKAP 893
Db 555 ELQGP-----RGSPNARSQP 569
RESULT 6
```



Nucleotide sequence alignment showing similarity between various sequences. The top section displays a query sequence (QY) aligned against several database entries (DB). The bottom section shows a multiple sequence alignment of nucleic acid sequences from different species.

**Query Match**

Best Local Similarity 4.9%; Score 244.5; DB 1; Length 706;  
Matches 99; Conservative 65; Mismatches 138; Indels 99; Gaps 15

**Sequence Alignment:**

The alignment consists of two main parts. The first part shows individual sequence matches with their respective accession numbers and scores. The second part shows a multiple sequence alignment of nucleic acid sequences, likely representing different isoforms or related genes across various species.

**Species and Accession Numbers:**

- AC P09405; Q61991;
- DT 01-MAR-1989 (Rel. 10, Created)
- DT 01-MAR-1989 (Rel. 10, Last sequence update)
- DT 15-JUN-2002 (Rel. 41, Last annotation update)
- DE Nucleolin (Protein C23).
- GN NCL OR NUC.
- OS Mus musculus (Mouse).
- OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
- OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
- OX NCBI\_TaxID=10090;
- RN [1]
- RN SEQUENCE FROM N.A.
- RP STRAIN=BALB/C;
- RC MEDLINE=88316930; PubMed=3137346;
- RA Bourbon H., Lapeyre B., Amalric F.;
- RT "Structure of the mouse nucleolin gene. The complete sequence reveals that each RNA binding domain is encoded by two independent exons.";
- RL J. Mol. Biol. 200:627-638(1988).
- RN [2]
- RN SEQUENCE OF 1-44 FROM N.A.
- RP MEDLINE=89121496; PubMed=2906027;
- RA Bourbon H.M., Prudhomme M., Amalric F.;
- RT "Sequence and structure of the nucleolin promoter in rodents: characterization of a strikingly conserved CpG island.";
- RL Gene 68:73-84(1988).
- RN [3]
- RN SEQUENCE OF 1-23.
- RP MEDLINE=91317840; PubMed=1860869;
- RX Pasternak M.S., Bleier K.J., McInerney T.N.;
- RA "Granzyme A binding to target cell proteins. Granzyme A binds to and cleaves nucleolin in vitro.";
- RL J. Biol. Chem. 266:14703-14708(1991).
- CC -!- FUNCTION: NUCLEOLIN IS THE MAJOR NUCLEOLAR PROTEIN OF GROWING EUKARYOTIC CELLS. IT IS FOUND ASSOCIATED WITH INTRANUCLEAR CHROMATIN AND PERIBIOSOMAL PARTICLES. IT INDUCES CHROMATIN CONDENSATION BY BINDING TO HISTONE H1. IT IS THOUGHT TO PLAY A ROLE IN PRE-RNA TRANSCRIPTION AND RIBOSOME ASSEMBLY.
- CC -!- SUBCELLULAR LOCATION: Nuclear; nucleolar.
- CC -!- SIMILARITY: CONTAINS 4 RNA RECOGNITION MOTIFS (RRM).
- CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL Outstation at the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).
- CC EMBL: X07699; CAA30538.1; .
- DR EMBL: M22089; AAA39841.1; .
- DR PTR: A2958; DMS.
- DR HSP: P09651; IHAL.
- DR MGD: MG1:97286; NCL.
- DR InterPro: IPRO00504; RNA\_rec\_mot.
- DR Pfam: PF00076; rtm; 4.
- DR SMART: SM00360; rrm; 4.
- DR PROSITE: PS50102; RRM; 4.
- KW Nuclear protein; Phosphorylation; Methylation; DNA-binding; Repeat; RNA-binding.
- KW INIT\_MET 0
- FT DOMAIN 142 169 ASP/GLU-RICH (ACIDIC).
- FT FT 189 214 ASP/GLU-RICH (ACIDIC).
- FT DOMAIN 240 272 ASP/GLU-RICH (ACIDIC).
- FT FT 308 384 RNA-BINDING (RRM) 1.
- FT FT 394 467 RNA-BINDING (RRM) 2.
- FT FT 486 560 RNA-BINDING (RRM) 3.
- FT FT 568 643 RNA-BINDING (RRM) 4.
- FT FT 645 696 ARG/GLY/PHE-RICH.
- FT FT 144 144 PHOSPHORYLATION (BY SIMILARITY).
- FT MOD\_RES 156 156 PHOSPHORYLATION (BY SIMILARITY).

[illegible]

RA Wiemann S., Weil B., Wellenreuther R., Gassenhuber J., Glassl S.,  
 RA Ansoerge W., Boecher M., Bloecker H., Bauersachs S., Blum H.,  
 RA Lauber J., Duesterhoeft A., Beyer A., Koehler K., Strack N.,  
 RA Mewes H.-W., Ottenwaelder B., Obermaier B., Tampe J., Heubner D.,  
 RA Wambutt R., Korn B., Klein M., Poustka A.,  
 RT "Towards a catalog of human genes and proteins: sequencing and  
 RT analysis of 500 novel complete protein coding human cDNAs",  
 RL Genome Res. 11:422-435(2001).  
 CC -1- SIMILARITY: CONTAINS 6 RNA RECOGNITION MOTIFS (RRM).  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL; AB014582; BAA31657.1; -  
 DR EMBL; AL117547; CAB55987.1; -  
 DR HSP; P09651; 1HA1.  
 DR SWISS-2DPAGE; O9Y4C8; HUMAN.  
 DR InterPro; IPR000504; RNA\_rec\_mot.  
 DR Pfam; PF00076; rrm; 6.  
 DR SMART; SM00360; RRM; 6.  
 DR PROSITE; PS0102; RRM; 6.  
 DR PROSITE; PS00030; RRM\_RNP\_1; 2.  
 KW RNA-binding; Repeat.  
 FT DOMAIN 2 79 RNA-BINDING (RRM) 1.  
 FT DOMAIN 294 369 RNA-BINDING (RRM) 2.  
 FT DOMAIN 402 480 RNA-BINDING (RRM) 3.  
 FT DOMAIN 587 659 RNA-BINDING (RRM) 4.  
 FT DOMAIN 730 811 RNA-BINDING (RRM) 5.  
 FT DOMAIN 832 912 RNA-BINDING (RRM) 6.  
 FT DOMAIN 225 232 POLY-GLU.  
 FT DOMAIN 391 396 POLY-GLU.  
 FT DOMAIN 714 725 POLY-GLU.  
 FT CONFLICT 141 141 R -> Q (IN REF. 2).  
 FT CONFLICT 259 259 R -> K (IN REF. 2).  
 FT CONFLICT 609 609 R -> H (IN REF. 2).  
 FT CONFLICT 623 623 T -> I (IN REF. 2).  
 FT CONFLICT 665 665 A -> T (IN REF. 2).  
 FT CONFLICT 921 921 R -> Q (IN REF. 2).  
 SQ SEQUENCE 960 AA; 107365 MW; 1D457D3B3D95377 CRC64;  
 Query Match 4.8%; Score 238.5; DB 1; Length 960;  
 Best Local Similarity 22.4%; Pred. No. 1.0e-05;  
 Matches 116; Conservative 80; Mismatches 173; Indels 149; Gaps 22;  
 QY 404 EKALNAGFIQATDY-VEIQAYLDYLRRYDFK-----QDSSK---ELELRA 447  
 DB 59 QKHFNKSFIDTSRITVECKSFSGDPAPRAWSKHAQKPSQKPPKDDSTTPEIKKDEKK 118  
 QY 448 AFTRALEYLKQVEERNEESGDPSCVIMQWRIEALCNMOKARELWDSIMTRGNKY 507  
 DB 119 KVAGOLEKKLKEDTE-----FQBFSLVHRRRAQAATWANDGLDAEPSKGSKP 165  
 QY 508 ANMWLEYNLERAGHDQHCRAKALHRAVQCTSDYPEHVCEVLLTMTERTGSLDWDIAVQ 567  
 DB 166 AS---DYLNFDSSGQE-----  
 QY 568 KTETRLARVNEQRMAKAEKAAALVQBEKEAEQKRAAEKALKKKKTRGPEKRADE 627  
 DB 189 -----LEEASL-----EPKAAVQKEL-SDMDYLKRWKVRAGSSSSSEEE 227  
 QY 628 DDEKE-----WGDDEEQPKRRRV---ENSIPAGETQNVVEAAGPAGK-----AAYDVE 676  
 DB 228 ESEDEAVHCDGSEAEEDSSATPVLOERDSRGAGQEGM-----PAGKKRPPARAETE 282  
 QY 677 PPSQKEKASLKDMPKVLHDSKDSITVFSNLPYSMQEPDTK--LRPLFEACGEVWQ 734  
 DB 283 KPANOKE-----PTTCH-----TVKLRGAPFNVTENKYMEEFLAPL-----RPVA 321

QY 735 IRPFSNRGDFRGYCYVEFKEEKSALQALEMDRKSVEGRPMFVSPCVKSKN-PDFKVF 793  
 DB 322 IRIRNAHGNKTYGIFVDFSEEBEVKQALKCNREYMGGR--YIE--VFREKNVPTTKGAP 377  
 QY 794 YSTS-----LEKHKLFIISGLPSCCKELEEETCKAHGTVKDLRL-VTN 835  
 DB 378 KNTTKSWOGRILGENEEEDLAESGRFLVRLNLPYTSTEDEKLFSKYGLPSLSELYPIDS 437  
 QY 836 RAGKPKGLAYVENESQASQAVMKMDMTIKENIKV 873  
 DB 438 LTKPKGFAITFPFPHAVKAYSEVDGQVFGQRLHV 475  
 RESULT 8  
 ELAV\_DROVI STANDARD; PRT; 519 AA.  
 AC P23241;  
 DT 01-NOV-1991 (Rel. 20, Created)  
 DT 01-NOV-1991 (Rel. 20, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Elav protein (Embryonic lethal abnormal visual protein).  
 GN ELAV.  
 OS Drosophila virilis (Fruit fly).  
 OC Eukaryota; Metazoa; Arthropoda; Mandibulata; Pancrustacea; Hexapoda;  
 OC Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;  
 OC Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.  
 OX NCBI\_TaxID=7244;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=91246165; PubMed=1903840;  
 RA Yao K.-M., White K.;  
 RT "Organizational analysis of elav gene and functional analysis of ELAV  
 RT protein of Drosophila melanogaster and Drosophila virilis";  
 RL Mol. Cell. Biol. 11:2994-3000(1991).  
 CC -1- FUNCTION: PROBABLY INVOLVED IN THE RNA METABOLISM OF NEURONS.  
 CC -1- SIMILARITY: CONTAINS 3 RNA RECOGNITION MOTIFS (RRM).  
 CC -1- SIMILARITY: BELONGS TO THE ELAV FAMILY OF RNP PROTEINS.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL; M61748; AAA28505.1; -  
 DR PIR; A40252; A40252.  
 DR HSP; P19339; 2SXL.  
 DR FlyBase; FBgn0013110; Dvir\elav.  
 DR InterPro; IPR002343; Hud\_Sxl\_RNA.  
 DR InterPro; IPR000504; RNA\_rec\_mot.  
 DR Pfam; PF00076; rrm; 3.  
 DR PRINTS; PRO0961; HUDSLRNA.  
 DR SMART; SM00360; RRM; 3.  
 DR PROSITE; PS0102; RRM; 3.  
 DR PROSITE; PS00030; RRM\_RNP\_1; 2.  
 KW RNA-binding; Repeat.  
 FT DOMAIN 23 164 ALA/GLN-RICH.  
 FT DOMAIN 185 276 RNA-BINDING (RRM) 1.  
 FT DOMAIN 284 365 RNA-BINDING (RRM) 2.  
 FT DOMAIN 438 516 RNA-BINDING (RRM) 3.  
 SQ SEQUENCE 519 AA; 55822 MW; B01AFIC0D601B0D9 CRC64;  
 Query Match 4.5%; Score 225; DB 1; Length 519;  
 Best Local Similarity 19.3%; Pred. No. 4.1e-05;  
 Matches 88; Conservative 103; Mismatches 154; Indels 110; Gaps 16;  
 QY 552 MERTEGSLDWDIAVOKETETRLA-----RVNEQRMKAAEKAALVQBEKEAEK 602  
 DB 50 LQQQQQQVQAILQVQQQQTQAVAAAAAAVTQQLQQQQQAVVQVQQQQQQQQQQQ 109  
 QY 603 RAAEKALKKKKK---IRGPEKRGADDEDEKEWGDDEEQPSKRRRRVENSIPAAGETQN 659









```
Db 1596 RQLEHYETELEDERKQRALAAAANKKLEGLDKOLELQADSAIKGRBEAIKQLRKLAQMK 1655
QY 623 RGADEDDKEKGWGD-----EQSPKRRRRVENSIPAAGETQNVVEAAGPAGKCAAVDVEP 677
Db 1656 DFQRELDARASRDEIFATSPKENEKRAKSLADLMQLE-----DLAAAEARAKQA-DLEK 1710
QY 678 PSQKQKAASLKMDPKVLHDSKDSITVPSNLPYSMQPDKLRLPLFACGEVVOIRP 737
Db 1711 EELAEELASSLSG--RNTLQD-EKRRLEARIQALEELEEEOGMEASDRVRKATLQAE 1767
QY 738 IFNRGDFRCYCYVEFEKEKSALQALEMDRKSVEGRPMFVSPCVDSKPNDFKVFYRSTIS 797
Db 1768 QLSN-----ELATERSTAQKNESARQQLERO-----1793
QY 798 LEKHLKLFISGLPFSCTKEELEECCKAHTGTVKDLRLVTRNAGKPKGLAYVEYE-----NES 852
Db 1794 -----NKLRSKLQVEGAVK-AKLASTVALEAKIAQLEEQVEQAREK 1837
QY 853 QASQAVNMKGDMGTIKENIKVAISNPPQKVPKPTRKAPGPMILLPOTYGARGKRTQ 912
Db 1838 QAATKSLKQKDKKLEKVLQV-----EDERKMAQYKEQAE-----KGNTK 1878
QY 913 LSLPLRALQ 921
Db 1879 VKQLKROLE 1887

RESULT 13
YL17_CAEEL
ID YL17_CAEEL STANDARD; PRT; 1130 AA.
AC Q1102;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical 131.5 kDa protein C02F12.7 in chromosome X.
GN C02F12.7.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Bristol N2;
RA Miller N.;
RL Submitted (NOV-1995) to the EMBL/GenBank/DBJ databases.
CC -!- SIMILARITY: WEAK, TO MYOSINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; U41545; AAK39135.1; -
CC WormPep; C02F12.7; CE03901.
CC Hypothetical protein; Coiled coil.
FT DOMAIN 121 779 COILED COIL (POTENTIAL).
FT DOMAIN 805 1061 COILED COIL (POTENTIAL).
SQ SEQUENCE 1130 AA; 131485 MW; 60FD2FE3D99FB09 CRC64;

Query Watch
Best Local Similarity 22.0%; Pred. No. 0.00035;
Matches 172; Conservative 123; Mismatches 261; Indels 226; Gaps 38;

QY 174 EKAVDYICPNLWIEYQYSGVIGGKGLKVRVSFVRALSSVGLHMTKGLALWE-----229
Db 135 EETVAEY-----ERQKYNVMTFS-----EYRERVAERERKLEAIEYSKIIALSEEVLG 183
QY 230 AYREFES-----ATVEAARLEKHVSL--FRQLAIPLYDMETATFAEYEEWSESDPIPEVS 281
Db 184 AKKDFEARKMSFOALQDREKEREKEQALEKLRKEHQEQVLEQRFSDTQLNLE-----237
```

```
QY 282 IQNYNKALQLEKYKPYEEALQAEAPRLAEYQAYIDFEMKIGDPARIQIOLIFRALV---338
Db 238 -QKYIIEIQRLLE-----ERKSLATEKERLGE-----TFENKLL--RRAQSLYETELTAAK 284
QY 339 -----ENCLVPDLMTIRYSQYLDR-----OLKVK-----DIVLSVHNRAIRNCPW 377
Db 285 MLYTKLEALRDHEEALKEELLARQDEFHDLQELQLOKRSREDLV-----SKN 335
QY 378 TVALNSRYLLAMER--HGVYDHOVISYTFEKNALNAGFIQATDYVEIQWALDYLRVRRYDEK 435
Db 336 DVT-----ALEKKLHNKEKEVQTLTKE-----LDQVKTTN-366
QY 436 QOSSKELEELRAAFTRALEYLVKOEVEERFESGDPSCVIMQNWARTEARLNCNMOKAREL 495
Db 367 -DKIRLTVTSEFAEYRKFKFOQEEE-----LRRKARLLTVVEAAKKEK 409
QY 496 WDSIMT--RGNAKYANMWLEYYNLERAHGDTQH-----CRKALHRAVOCTSD 540
Db 410 LESVISDLQVEVKALKNKVEFLEKERENLOSQSESOTQLOSSQVDALEAVLHVSVTKEKET 469
QY 541 YPEHVCVLLTMTERTEGSLEDWDIAVOKT-ETRLARVNEQ--RMKAAEKEAALVQOEFEK 597
Db 470 TREHY-EGLLDKEROQA--ESREHAMKKEFCKLNELEEOYTSLSKEEESARL-----DK 522
QY 598 AEQRKRAAEKALKKKKIRGPEKRGADDEDEKEMGDDDEEOPSKRRRVRVENSIPAAGET 657
Db 523 DELREASEIEIQALRTEKLSILAAEIRVLTKQTEDEODDITQOLAKIVEDTSQITLLEE 582
QY 658 QNVEAAGPAGCAAVDVEPPSKQKQKAAASLRDMPKVLHDSKDSITVFSNLPYSMOE 717
Db 583 YRERI-----TGKDA--EILNLRKQLEKEISHTEDNRLLHENTQKELE-----AHKE 628
QY 718 PTKLRPLFEACEGVVQIRPIFSNRGDFRCYCYVEFEKEKSALQALEMDRKS-770
Db 629 THTETVRVLEA--EIDQKSAFENEQYEG-----KEKSAKIRELEAQNKLLSMEKV 679
QY 771 ----EGRPMFVSPVDK-----SKNPDFKVFYRSTSLKHLKLFISGLPFSCTKEELE 818
Db 680 KHVAENLEAFTS--DKONLLEESKKNKNIHLKQETIAQLNEKI-----STKE---725
QY 819 EICKAHGTVKDLRLVTRNAGKPKGLAYVEYENESQASQAV---MKMDGM-----TIKENI 870
Db 726 -----TEKQSEL-----EKTIAQLEIDNSSKSDQIEKLHLRVNMDLQMGTKIDEL 771
QY 871 IK 872
Db 772 VK 773
```

## RESULT 14

## GAR2\_SCHPO

ID GAR2\_SCHPO STANDARD; PRT; 500 AA.

AC P41891; O13707;

DT 01-NOV-1995 (Rel. 32, Created)

DT 30-MAY-2000 (Rel. 39, Last sequence update)

DT 15-JUN-2002 (Rel. 41, Last annotation update)

DE Protein gar2

GN GAR2 OR SPAC140.02.

OS Schizosaccharomyces pombe (Fission yeast).

OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;

OC Schizosaccharomycetales; Schizosaccharomycetaceae;

OC Schizosaccharomycetes.

OX NCBI\_TaxID=4896;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=972;

RX MEDLINE=95319932; PubMed=7596817;

RA Gulli M.-P., Girard J.-P., Zabetakis D., Lapeyre B., Melese T.,

RA Caizergues-Ferrer M.;

RT "gar2 is a nucleolar protein from Schizosaccharomyces pombe required

RT for 18S rRNA and 40S ribosomal subunit accumulation.";

RL Nucleic Acids Res. 23:1912-1918(1995).



[illegible]

DB 231 SSSDEKKRKAAPPASEERPAKITRPSQSDSNCTCTVFVGRLSWNV--DOWLGEFEYGT 288  
QY 732 VQIPIFISNR-GDPRGYCYVEFEKEKSAQALEMD-RKSVEGRPMFVSPCVDKSKNPDF 789  
DB 289 IVGARVMDGSGRSKGYVDFTPEPAKAAVAANGTKEDGMVNDLSNRPANQP 348  
QY 790 KV-----FRYSTSLKHLKFLISGLPFCSTCKEELEETCKAHTGTVKDLRLVTN-RAGPKG 842  
DB 349 YAAQRAGNFQDLSEPSDTFVGNLSFNATDEDLSTAFGGCGDIQSIRLPTDPSGRLKG 408  
QY 843 LAYVEYENESQASQVAKMDGMTIKENLIKVAISNP 878  
DB 409 FGVTFTSDIDSAAKCV-EMNGHFIAGRPCRLDFTSTP 443

RESULT 15  
ELAV DROME  
ID ELAV DROME STANDARD; PRT; 483 AA.  
AC P16914; Q9V3F6;  
DT 01-AUG-1990 (Rel. 15, Created)  
DT 01-AUG-1990 (Rel. 15, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Elav protein (Embryonic lethal abnormal visual protein).  
GN ELAV OR EG:65Fl.2 OR CG4262.  
OS Drosophila melanogaster (Fruit fly).  
OC Eukaryota; Metazoa; Arthropoda; Mandibulata; Pancrustacea; Hexapoda;  
OC Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;  
OC Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.  
OX NCBI\_TaxID=7227;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=Canton-S;  
RX MEDLINE=89072740; PubMed=3144044;  
RA Robinow S., Campos A.R., Yao K.-M., White K.;  
RT "The elav gene product of Drosophila, required in neurons, has three  
RNP consensus motifs.";  
RL Science 242:1570-1572(1988).  
RN [2]  
RP ERRATUM.  
RA Robinow S., Campos A.R., Yao K.-M., White K.;  
RL Science 243:12-12(1989).  
RN [3]  
RP SEQUENCE FROM N.A.  
RC STRAIN=Berkeley;  
RX MEDLINE=20196006; PubMed=10731132;  
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,  
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,  
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,  
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,  
RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champagne M., Pfeiffer B.D.,  
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,  
RA Allred J.F., Agbayani A., An H.-J., Andrews-Pfannkuch C., Baldwin D.,  
RA Baller R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,  
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,  
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brothier P.,  
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,  
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,  
RA de Fabros B., Deicher A., Deng Z., Hays A.D., Dew I., Dietz S.M.,  
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,  
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,  
RA Flocker C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,  
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,  
RA Harris N.L., Harvey D., Helman T.J., Hernandez J.R., Houck J.,  
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,  
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,  
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,  
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,  
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,  
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,  
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,  
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,  
RA Palazzolo M., Pittman G.S., Pan S., Pollard R.D., Puri V., Reese M.G.,

[2]  
RC STRAIN=972;  
MEDLINE=21848401; PubMed=11859360;  
RA Wood V., Gilliam R., Rajandream M.A., Lyne M., Lyne R., Stewart A.,  
RA Sgouros J., Peat N., Hayles J., Baker S., Basham D., Bowman S.,  
RA Brooks K., Brown D., Brown S., Chillingworth T., Churcher C.M.,  
RA Collins M., Connor R., Cronin A., Davis P., Feltwell T., Fraser A.,  
RA Holtroyd S., Goble A., Hamlin N., Harris D., Hidalgo J., Hodgson G.,  
RA Holtenby S., Howarth S., Huckle E.J., Hunt S., Jagels K.,  
RA James K., Jones L., Jones M., Leather S., McDonald S., McLean J.,  
RA Mooney P., Moule S., Mungall K., Murphy L., Niblett D., Odell C.,  
RA Oliver K., O'Neil S., Pearson D., Quail M.A., Rabinowitsch E.,  
RA Rutherford K., Rutter S., Saunders D., Seeger K., Sharp S.,  
RA Skelton J., Simmonds M., Squares R., Squares S., Stevens K.,  
RA Taylor K., Taylor R.G., Tivey A., Walsh S.V., Warren T., Whitehead S.,  
RA Woodward J., Volkart G., Aert R., Robben J., Grymonprez B.,  
RA Weltyens I., Vanstreels E., Rieger M., Schaefer M., Mueller-Auer S.,  
RA Gabel C., Fuchs M., Fritze C., Holzer E., Moestl D., Hilbert H.,  
RA Borzym K., Langer I., Beck A., Lehrach H., Reinhardt R., Pohl T.M.,  
RA Eger P., Zimmermann W., Wedler H., Wambutt R., Purnelle B.,  
RA Goffeau A., Cadieu E., Dreano S., Gloux S., Lelaure V., Mottier S.,  
RA Galibert F., Aves S.-J., Xiang Z., Hunt C., Moore K., Hurst S.M.,  
RA Lucas M., Rochet M., Gaillardin C., Tallada V.A., Garzon A., Thode G.,  
RA Daga R.R., Cruzado L., Jimenez J., Sanchez M., del Rey F., Benito J.,  
RA Dominguez A., Revuelta J.L., Moreno S., Armstrong J., Forsburg S.B.,  
RA Cerutti L., Lowe T., McCombie W.R., Paulsen I., Potashkin J.,  
RA Shpakovski G.V., Ussery D., Barrell B.G., Nurse P.;  
RT "The genome sequence of Schizosaccharomyces pombe.";  
RL Nature 415:871-880(2002).  
CC -!- FUNCTION: HELPS THE ASSEMBLY OF PRE-RIBOSOMAL PARTICLES  
CC CONTAINING 18S RNA.  
CC -!- SUBCELLULAR LOCATION: Nuclear; nucleolar.  
CC -!- SIMILARITY: BELONGS TO THE GAR FAMILY.  
CC -!- SIMILARITY: CONTAINS 2 RNA RECOGNITION MOTIFS (RRM).  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL; Z48166; CA88179.1; .  
CC EMBL; AL163191; CAB86413.1; .  
CC HSP: P11940; 1CVI  
CC InterPro: IPR000504; RNA\_rec\_mot.  
CC HSP: P11940; 1CVI  
CC Pram: PF00076; rtm; 2.  
CC SMART; SM00360; RRM; 2.  
CC PROSITE; PS0102; RRM; 2.  
CC PROSITE; PS00030; RRM\_RNP\_1; 2.  
CC Ribosome biogenesis; RNA-binding; Nuclear protein; Repeat;  
CC RNA processing.  
CC FT DOMAIN 81 236 GLU/SER-RICH.  
FT DOMAIN 263 341 RNA-BINDING (RRM) 1.  
FT DOMAIN 366 443 RNA-BINDING (RRM) 2.  
FT CONFLICT 339 339 S -> P (IN REF. 1).  
KW SEQUENCE 500 AA; 52987 MW; 9D37FAD0C5161A0B CRC64;  
Query Match 4.2%; Score 211.5; DB 1; Length 500;  
Best Local Similarity 22.6%; Pred. No. 0.00021;  
Matches 76; Conservative 76; Mismatches 147; Indels 37; Gaps 12;  
QY 567 QKTEFLARVNEQRKAAEKAALVQEEKA-----PQRKARAEKALKKKKIRGPEK 622  
DB 121 ESEEEVYVTKTEKSSSESSSESEEEAAVYKIEKKSSSDSS-----ES 170  
QY 623 RGADEDEKEMGDDEEO---PSKRRRVVENSIPAAGETONVYAAGPAGKC-RAVDVEPP 678  
DB 171 SSSSESSSSSESEEEVEVKEETKEKGGSSSSSSSSSSSSSSSSSSSSSSSSSE 230  
QY 679 S-----KQKKAASLRKMPKVLHPDSSKDS-----ITVFVSNLPYSMQEPTDKLRPLFEACGE 731

